João de Fernandes Teixeira

MENTE, CÉREBRO E COGNIÇÃO

EDITORA VOZES

Petrópolis

© 2000, Editora Vozes Ltda.
Rua Frei Luís, 100
25689-900 Petrópolis, RJ
Internet: http://www.vozes.com.br
Brasil

Todos os direitos reservados. Nenhuma parte desta obra poderá ser reproduzida ou transmitida por qualquer forma e/ou quaisquer meios (eletrônico ou mecânico, incluindo fotocópia e gravação) ou arquivada em qualquer sistema ou banco de dados sem permissão escrita da Editora.

Diretor editorial
Frei Antônio Moser

Editores
Aline dos Santos Carneiro
José Maria da Silva
Lídio Peretti
Marilac Loraine Oleniki

Secretário executivo
João Batista Kreuch

Editoração e org. literária: Enio P. Giachini
Projeto gráfico: AG.SR Desenv. Gráfico
Capa: Marta Braiman

ISBN 978-85-326-2349-2

Editado conforme o novo acordo ortográfico.

Este livro foi composto e impresso pela Editora Vozes Ltda.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Teixeira, João de Fernandes

Mente, cérebro e cognição / João de Fernandes Teixeira.
4. ed. – Petrópolis, RJ : Vozes, 2011.

Bibliografia.

ISBN 978-85-326-2349-2

1. Cérebro 2. Ciência cognitiva 3. Filosofia da mente I. Título.

00-1552 CDD-128.2

Índice para catálogo sistemático:

1. Filosofia da mente e ciência cognitiva 128.2
2. Mente e cérebro : Filosofia 128.2

Este livro é dedicado a Lydia Braga Teixeira, in memoriam.

Agradecimentos

A maior parte deste livro foi escrita em 1998, quando realizei um estágio de pós-doutoramento no Centro de Estudos Cognitivos da Tufts University, Boston, Estados Unidos. Agradeço ao diretor do Centro, Professor Daniel Dennett, pela oportunidade e à Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo apoio financeiro. Cabe também agradecer aos meus colegas do curso de pós-graduação em filosofia da Universidade Federal de São Carlos que, nesse período, assumiram minhas atividades docentes, tornando possível meu estágio no exterior.

Meus ex-orientandos Marcos Romano Bicalho e Saulo de Freitas Araujo contribuíram na elaboração de partes dos capítulos II e V, respectivamente, e a eles fica expressa minha gratidão. A Saulo devo também inúmeras críticas e sugestões que surgiram de sua leitura das versões preliminares deste livro.

Sumário

Introdução, 11

CAPÍTULO I: MENTE E CÉREBRO, 15

CAPÍTULO II: A HERANÇA CARTESIANA, 29

Uma incursão pelas "Meditações" de Descartes, 31
Acesso privilegiado e não espacialidade, 34
A teoria dos autômatos, 36
A filosofia pós-cartesiana: uma história abreviada, 40
O século XX: teoria da relatividade, 44
O século XX: a neurociência, 51

CAPÍTULO III: MATERIALISMO E TEORIAS DA IDENTIDADE, 65

Algumas distinções técnicas e terminológicas, 66
Teorias da identidade, 69
Teorias reducionistas, 71
Emergentismo e superveniência, 79
O futuro do materialismo: a noção de matéria, 85

CAPÍTULO IV: AS VARIEDADES DO DUALISMO, 89

Thomas Nagel, 93
David Chalmers, 98
O futuro do dualismo, 106

CAPÍTULO V: DESFAZENDO A IDEIA DE MENTE, 107

Gilbert Ryle, 111
O materialismo eliminativo dos Churchlands, 115
O futuro do materialismo eliminativo, 119

CAPÍTULO VI: FUNCIONALISMO E MENTES ARTIFICIAIS, 123

A linguagem do pensamento e a teoria representacional da mente, 128
As objeções ao funcionalismo, 132
Algumas respostas, 138
Mais algumas respostas, 141
A teoria dos sistemas intencionais, 143
Mente, cérebro e algoritmos de compressão, 148

CAPÍTULO VII: TEORIAS DA CONSCIÊNCIA, 153
Uma primeira incursão: Dennett, Calvin e Baars, 160
A redescoberta do cérebro, 165
O fim do funcionalismo?, 170
O futuro da neurociência cognitiva, 173
Conclusão, 177
Bibliografia comentada, 185

Nesse dia – cuido que por volta de 2222 – o paradoxo despirá as asas para vestir a japona de uma verdade comum. [...] As filosofias queimarão todas as doutrinas anteriores, ainda as mais definitivas, e abraçarão essa psicologia nova, única verdadeira, e tudo estará acabado. [...] Olha bem que é a cabeça do cônego. Temos à escolha um ou outro dos hemisférios cerebrais; mas vamos por este, que é onde nascem os substantivos. Os adjetivos nascem no da esquerda. Descoberta minha, que, ainda assim, não é a principal, mas a base dela, como se vai ver. Sim, meu senhor, os adjetivos nascem de um lado e os substantivos de outro [...].

(Machado de Assis, ***O cônego ou a metafísica do estilo****).*

Introdução

Este livro fala de mentes, cérebros e de sua relação. Seu principal objetivo é fornecer ao leitor universitário uma introdução à filosofia da mente e suas relações com a ciência cognitiva – esses dois grandes continentes que ainda se mantêm, em parte, desconhecidos do público brasileiro.

A filosofia da mente consolidou-se no século XX, tendo resultado de uma jornada milenar que a torna uma disciplina de história curta, mas de passado longo. Falar de mentes já foi privilégio de místicos, textos sagrados e de filósofos refugiados em teorias herméticas. Nos últimos anos esse passa a ser um tema abordado pela racionalidade profana de nossa ciência. Na filosofia da mente contemporânea aliam-se ciência e reflexão filosófica, numa combinação imposta por se reconhecer a necessidade de uma investigação interdisciplinar.

A mente deixou de ser algo exclusivo dos seres humanos. Desde a década de 1940 passamos a atribuir mentes e inteligência a máquinas e outros dispositivos artificiais. Desenvolveu-se uma *tecnologia do mental*, da qual resultou uma aproximação crescente entre a psicologia, ciência da computação e a engenharia. Desse projeto interdisciplinar surgiu a inteligência artificial e, posteriormente, a ciência cognitiva. A inteligência artificial teve um triunfo efêmero na década de 1970, o qual, entretanto, foi suficiente para mostrar que muitas atividades consideradas exclusivas dos seres humanos, como, por exemplo, jogar xadrez, fazer cálculos de engenharia etc. poderiam ser feitas por computadores bem programados.

A replicação tecnológica da inteligência e das atividades mentais vem tendo consequências profundas sobre o modo como concebemos a relação entre mente e cérebro. Ela sugere que aquilo que chamamos de "mente" talvez não seja mais do que um tipo específico de arranjo material, feito a partir de peças de silício.

Paralelamente à revolução computacional e seus desdobramentos mais recentes, ingressamos, a partir dos anos de 1990, na "década do cérebro". Nela se esperava que o desenvolvimento da neurociência, aliado aos progressos de outras disciplinas como a genética e a biologia molecular, pudesse finalmente desvendar a natureza da consciência humana – que alguns já declararam ser o último mistério ainda não resolvido pela ciência. A década do cérebro já terminou, grandes avanços foram alcançados, mas a natureza da consciência ainda continua sendo um mistério. Desta década ficaram, entretanto, marcas profundas: nela, mais do que em qualquer época, tentou-se tornar a ciência da mente uma ciência do cérebro.

Nossas alegrias, tristezas, pensamentos e outros estados subjetivos nada mais seriam do que o resultado da atividade de alguns grupos de neurônios do nosso cérebro. Podemos controlar quimicamente nossa ansiedade e nossa angústia, o que viria a con-

firmar que essas nada mais seriam do que desconfortáveis ilusões subjetivas. Mas, ao dissipar ilusões subjetivas a neurociência parece caminhar em direção a dissipar o próprio conceito de mente. Ao livrar-se da ideia de mente, ela estaria jogando fora o bebê junto com a água do banho.

Entre a tecnologia do mental e a neurociência haveria ainda algo que teria se tornado uma terra de ninguém: a psicologia, que, durante muitos anos, tinha permanecido como reduto privilegiado daqueles que queriam falar sobre mentes. Para os neurocientistas e para os engenheiros do mental a psicologia está vivendo seus últimos dias, estando fadada a desaparecer num futuro próximo, da mesma maneira que a alquimia foi substituída pela química. "Mente" estaria se tornando um conceito obsoleto.

Desde seu aparecimento, há pouco mais de 150 anos, a psicologia tem enfrentado uma desorganização teórica profunda. Nas últimas décadas ela se tornou particularmente aguda, ao ponto de filósofos como o austríaco Ludwig Wittgenstein a ironizarem com sentenças bombásticas como "Na psicologia há métodos experimentais e confusão conceitual"[1]. Os psicólogos nunca teriam realmente sabido do que estavam falando ao se referirem a mentes. Desse pântano não parecem ter escapado os tecnólogos da mente nem tampouco os neurocientistas. Quando lemos atentamente seus trabalhos ou visitamos seus laboratórios frequentemente ficamos com a impressão de que muitas vezes aqueles que se envolvem nesse tipo de empreendimento não sabem exatamente o que estão fazendo. É como se tivessem um excelente navio, com uma tripulação altamente qualificada sem, contudo, saber de onde se partiu e para onde se está navegando, correndo o risco de algumas vezes tomar uma ilhota por um continente ou cometendo o erro oposto. O conceito de mente ainda parece constituir o grande ponto cego da investigação científica.

Na contramão desse movimento científico encontramos uma forte reação à invasão progressiva da ciência nos últimos redutos do mundo da mente: o misticismo. Com ele busca-se reencontrar algum tipo de reencantamento do mundo, opondo-se à dessacralização e ao processo de naturalização da mente, ou seja, a redução dessa ao substrato químico e biológico do cérebro. Do misticismo desliza-se facilmente para a mistificação. A psicologia não ficou imune a esse tipo de movimento, encontrando-se hoje invadida por doutrinas exóticas dos mais variados tipos.

O desafio que enfrentamos é, então, o de desenvolver um conceito de mente e de sua relação com o cérebro que acomode a possibilidade de uma investigação científica interdisciplinar; uma investigação que concilie nossa própria descrição como cérebros e organismos com nossa descrição como pessoas dotadas de mentes. Não poderíamos resolver esse problema decretando unilateralmente o fim da ideia de mente ou sustentando que essa passará para a lista dos conceitos científicos obsoletos, da mesma maneira que o "flogisto" foi substituído pelo oxigênio.

1. Wittgenstein (1951), parágrafo 360.

O problema da natureza da mente e de sua relação com o cérebro ainda extrapola o âmbito da investigação científica de que dispomos. Mas se a ciência não pode resolver essa outra ordem de problemas – problemas filosóficos ou conceituais – não podemos, tampouco, virar as costas para ela. A filosofia tradicional tem se debatido, durante séculos, com a questão de saber se o cérebro *produz* a mente ou se ele apenas a *manifesta*, sendo apenas um complexo e misterioso hospedeiro biológico. Da filosofia tradicional teríamos legado apenas a aridez metafísica; da neurociência e da engenharia do mental a excessiva ingenuidade filosófica de alguns cientistas. Cabe à filosofia da mente buscar uma terceira margem do rio ou uma perspectiva da qual possamos, quando falamos de mentes e de cérebros, distinguir entre cavaleiros e moinhos de vento.

Não pretendemos, com este livro, oferecer uma exposição completa de todas as teorias e de tudo o que podemos dizer acerca das relações entre mente e cérebro. Este é um texto introdutório, que deve ser visto como o resultado de um rápido voo de reconhecimento sobre uma grande cidade, um voo no qual se descortinam várias perspectivas sem que, entretanto, possamos integrá-las ao ponto de podermos desenhar um mapa completo[2].

Iniciamos este trabalho mostrando em que sentido o problema das relações entre mente e cérebro constitui um problema conceitual ou filosófico (capítulo I). Buscamos suas raízes históricas (capítulo II), para então percorrer as principais tentativas de solução propostas pela filosofia da mente no século XX: as teorias que propõem que mente e cérebro são a mesma coisa, como o materialismo contemporâneo (capítulo III); as teorias que se ocupam de negar que mente e cérebro sejam a mesma coisa e que ressuscitam o dualismo (capítulo IV) e a proposta de desfazer o conceito de mente, tratando a questão da relação entre mente e cérebro como um pseudoproblema (capítulo V).

No capítulo VI expomos o chamado "modelo computacional da mente" e com ele algumas tentativas recentes e originais de resolver o problema mente-cérebro a partir da moderna tecnologia de construção de computadores e robôs. Essas tentativas consolidaram-se a partir do final dos anos de 1960 e foram batizadas com o nome genérico de *funcionalismo*. Para os funcionalistas, a mente é um tipo específico de organização que ocorre no cérebro humano, mas que pode ocorrer também num computador ou em algum outro tipo de dispositivo artificial.

O sétimo e último capítulo envereda por uma apresentação das teorias contemporâneas da natureza da consciência. Esse tema impressiona não apenas pela enorme proliferação teórica que tem produzido nos últimos anos, como também pela perplexidade que tem suscitado. Busca-se uma teoria geral da consciência e da subjetividade, seja através de simulações computacionais, seja através do estudo do cérebro. A busca incessante dos correlatos neurais da consciência leva-nos para um outro tipo de dis-

2. Uso esta mesma metáfora em Teixeira (1996a).

cussão que tem ocupado a filosofia da mente contemporânea: avaliar o significado das novas técnicas de mapeamento cerebral através das tecnologias de neuroimagem, que se consolidaram nas últimas décadas e que apontariam para a possibilidade de explicar a natureza dos fenômenos mentais a partir de propriedades específicas do cérebro humano.

Em todo esse percurso procuramos nunca perder de vista que não estamos escrevendo para filósofos acadêmicos, mas para todos aqueles que, direta ou indiretamente, ocupam-se da natureza da mente nas suas diversas profissões. Contudo, ao escrever um livro que, se não é totalmente filosófico, pelo menos preserva um tom filosófico, não seria possível deixar de tomar partido ou sustentar alguma posição específica. Inclino-me em favor da teoria dos sistemas intencionais, formulada pelo filósofo norte-americano Daniel Dennett e que apresento na segunda parte do capítulo VI. Não me ocupo, entretanto, em fazer nenhuma defesa explícita desse tipo de posição, restringindo-me à tarefa de oferecer subsídios para que o leitor possa tirar suas próprias conclusões. Somente nas últimas páginas deste livro defendo abertamente esse ponto de vista, mostrando em que sentido a teoria dos sistemas intencionais permite uma conciliação entre, de um lado, a preservação da ideia de mente e, de outro, sua descrição biológica proposta pela investigação neurocientífica. Sugiro que o conceito de "mente" é um conceito operacional, tão imprescindível para a compreensão de nossos fenômenos mentais e dos fenômenos mentais de outros seres humanos quanto o conceito de "centro de gravidade" o é para a física. Essa seria a terceira margem do rio.

Numa tentativa de contornar as dificuldades que o leitor possa encontrar e, tendo em vista também a possibilidade de utilizar este trabalho como livro universitário, optamos por uma linguagem coloquial e didática na maior parte do texto. Algumas subsecções, nas quais foi inevitável o uso de um jargão mais especializado ou que, pelo seu caráter técnico, tornaram-se mais difíceis, estão marcadas com um asterisco (*). Nesses casos, aconselhamos o leitor que simplesmente as pule; a leitura não sequencial deste livro foi prevista como uma possibilidade que não prejudicará excessivamente a sua compreensão.

Ao final de cada capítulo o leitor encontrará, também, sugestões de leituras que poderão ser usadas para consulta ou para se aprofundar em alguns assuntos que o interessem mais. A mesma estratégia foi usada para a elaboração de uma bibliografia comentada que acrescentamos ao final deste texto. Nela procuramos listar não apenas os títulos citados e indicados ao longo desta obra, como também livros mais recentes e importantes para que o leitor possa ter acesso a informações específicas e mais atualizadas. Não poderíamos ignorar a enorme quantidade de textos, informações e bibliografias sobre filosofia da mente e ciência cognitiva que proliferam na internet – algumas muito úteis e outras de qualidade duvidosa. Embora endereços de "sites" eletrônicos possam mudar de tempos em tempos, indicamos na penúltima seção algumas URLs que contêm informações importantes e textos de boa qualidade que poderão ajudar o leitor a complementar sua jornada.

CAPÍTULO I

MENTE E CÉREBRO

A primeira questão colocada pela filosofia da mente é: serão mente e corpo a mesma coisa? Será o pensamento apenas um produto do meu cérebro – que produziria pensamentos da mesma forma que meu pâncreas produz insulina? Qual é a natureza dos fenômenos mentais?

Essa não é apenas uma primeira questão numa ordem de indagações. Trata-se da pergunta mais importante a ser respondida pela filosofia da mente – o problema fundamental que dá origem a quase todos os temas tratados por essa disciplina.

Eu posso fechar meus olhos e, numa fração de segundos, pensar em estrelas coloridas cintilando num céu azul escuro. Estrelas que nem sequer sei se existem, e que talvez estejam a muitos anos-luz de distância. Eu posso imaginar uma vaca amarela ou então dizer que estou sentindo muito calor. Entretanto, se alguém pudesse abrir o meu cérebro e examiná-lo com o mais aperfeiçoado instrumento de observação de que a ciência dispõe, não veria estrelas coloridas nem uma vaca amarela. Veria apenas uma massa cinzenta, cheia de células ligadas entre si.

Essas células são chamadas de *neurônios*, verdadeiras unidades do sistema nervoso cuja existência foi finalmente provada somente há cerca de um século com o trabalho de S. Ramón y Cajal. Até então muitos achavam que o sistema nervoso era um conjunto de vias contínuas, subdivididas em minúsculos filamentos. Os neurônios têm diversas formas e tamanhos, tendo, todos, entretanto, uma região destinada a fazer contato com outros neurônios, os chamados dendritos. O corpo da célula, o soma, contém um núcleo e outras estruturas, como os mitocôndrias, que participam dos aspectos metabólicos da atividade dos neurônios. Há também uma outra conexão de um neurônio com outros, mais longa e através da qual se movimenta o impulso nervoso. Essa conexão é chamada de axônio. Cada região do neurônio revela propriedades elétricas, mas os impulsos geralmente ocorrem, na maioria das vezes, no axônio.

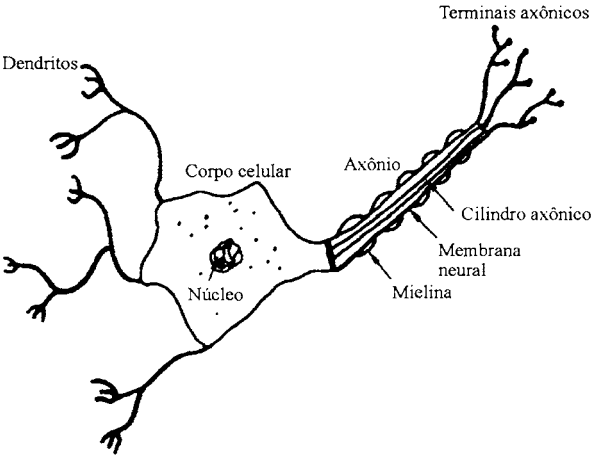

Fig. 1.1 – Esquema geral de um neurônio.

Desde o aparecimento dos trabalhos de Ramón y Cajal, nenhuma outra disciplina se desenvolveu tanto neste século XX quanto a neurociência. Dispomos hoje de um conhecimento bastante preciso do funcionamento cerebral e de suas unidades básicas, bem como das reações químicas que nele ocorrem. Sabemos que o cérebro é uma máquina complexa resultante da reunião de elementos fundamentais: o neurônio ou unidade básica, as sinapses ou conexões entre os neurônios e as ligações químicas que ali ocorrem, através de neurotransmissores e receptores. Essas combinações o tornam uma máquina extremamente poderosa, na medida em que são capazes de gerar configurações e arranjos variados num número astronômico.

Contudo, o grande desafio que a neurociência ainda enfrenta é a dificuldade (ou será uma impossibilidade?) de relacionar o que ocorre no cérebro com aquilo que ocorre na mente, ou seja, de encontrar algum tipo de *tradução* entre sinais elétricos das células cerebrais e aquilo que percebo ou sinto como sendo meus pensamentos. A observação da atividade elétrica de meu cérebro não permite saber se estou pensando em estrelas coloridas ou numa vaca amarela. Alguém poderia até inferir – de algum tipo de observação do que ocorre no meu cérebro – que estou sentindo calor, mas não saberia dizer se o calor que eu sinto é maior ou menor do que o calor que o cientista, ao observar meu cérebro, estaria sentindo.

Se ninguém pode observar esses fenômenos ocorrendo em mim e se ninguém os encontra em meu cérebro, então posso formular duas perguntas: Onde eles estarão ocorrendo? E o que serão eles se – pelo menos inicialmente – não posso supor que sejam objetos como quaisquer outros que se apresentam diante de mim, como parte da natureza?

Essas duas questões estão na origem da determinação daquilo que chamamos de subjetividade. Estrelas coloridas e cintilantes, bem como vacas amarelas, existem *para mim*, pelo menos momentaneamente. Se ninguém mais pode observá-las, posso

O problema ontológico é, na verdade, milenar. Foi Platão quem o propôs pela primeira vez, ao inventar (ou descobrir?) as *ideias*. Ele propôs que nossos conteúdos mentais podem ser abstraídos e individuados das mentes que os pensam. As ideias ou conteúdos mentais apenas *ocorrem* nas mentes, tendo uma realidade independente dessas. Os pensamentos podem ser abstraídos dos atos de pensá-los, formando um verdadeiro mundo à parte. Pensar significa apenas aceder a esse mundo, incorporar as ideias ou pensamentos aos nossos atos de pensar, tornando-nos apenas veículos momentâneos desse mundo das ideias. Perceberíamos a realidade independente desse mundo das ideias quando pensamos, por exemplo, nas verdades da matemática e quando concebemos, por exemplo, que "2 + 2 = 4". A realidade de 2 + 2 = 4 ou de uma demonstração geométrica qualquer seriam *sempre* verdadeiras independentemente de elas ocorrerem ou não na minha mente. Uma verdade geométrica como, por exemplo, "a soma dos ângulos de um triângulo será sempre 180 graus" independe não só de eu pensá-la aqui e agora como também de qualquer triângulo ou exemplo de triângulo que possa existir no mundo e que eu possa estar vendo na minha frente. Isto levou Platão a propor que o mundo das ideias é o único mundo verdadeiro, o mundo imutável. O mundo que percebemos através dos nossos sentidos ou que pensamos momentaneamente seria apenas uma cópia desse mundo verdadeiro. Com isto, Platão dividiu a realidade em duas partes: a do mundo sensível e a do mundo inteligível.

Platão inaugurou a *dualidade de realidades* ou o *dualismo ontológico*. Ao inventar o mundo das ideias espalhou a discórdia entre os filósofos, que nunca mais chegaram a um consenso sobre o que existe ou não, se o mundo é aquilo que vemos ou se existe algo para além daquilo que os nossos sentidos nos mostram. Nem mesmo um consenso sobre quantas e quais ideias existiriam nesse mundo jamais foi alcançado. Existiria apenas uma ideia para cada tipo de objeto sensível, ou cópia dessa ideia, ou várias ideias correspondentes a várias cópias desse objeto que poderiam estar no mundo sensível? E quando penso no cavalo alado, terá ele uma existência real e independente nesse mundo das ideias? Essa foi a primeira e a maior diafonia da história da filosofia; tudo o que se seguiu depois foi uma tentativa de superá-la.

Quando refletimos sobre o problema da relação entre mente e cérebro podemos perceber como essa diafonia se manifesta. Serão nossos pensamentos parte do mundo das ideias? Serão nossos cérebros apenas veículo desses pensamentos, transmissores de algum tipo de versão sensível desse mundo das ideias, produtores apenas de cópias manifestas dessa realidade independente? Podemos dar um passo a mais e perguntar também se "mente" não seria uma ideia no sentido platônico, ou seja, uma realidade única, imutável e eterna, da mesma maneira que as verdades matemáticas e geométricas. Mas sobre isto os filósofos discordaram. Alguns acharam que a ideia de "mente" não seria nada além de um artifício de imaginação, como, por exemplo, o cavalo alado. Outros acharam que mentes existiriam nesse mundo à parte, sem, contudo, conseguir explicar como essas poderiam se relacionar com o mundo sensível. De uma forma ou de outra, estava formada a diafonia, a cisão fundamental ou o problema ontológico, do qual o problema da relação entre mente e cérebro seria apenas um caso particular, embora de importância decisiva. Dualistas e monistas jamais puderam se reconciliar e a

história da filosofia da mente repete o longo comentário e tentativa de reconciliação entre os dois mundos de Platão.

Muitos de nós somos platonistas sem o saber. O platonismo é uma espécie de filosofia espontânea dos matemáticos, de engenheiros e de computólogos, que acreditam que a mente é uma estrutura matemática abstrata que poderia ser reproduzida por um programa de computador. A mente seria independente do cérebro, e, da mesma maneira que um teorema matemático seria independente do cérebro de alguém que o descobriu, o programa de computador seria independente do computador onde ele é rodado. Esse programa de computador seria portável, no sentido computacional do termo, ou seja, poderia ser transportado e instalado em qualquer tipo de hardware. Instalaríamos mente e consciência num computador da mesma maneira que nele instalamos – através de um CD-Rom – o Windows 98. Teríamos um ícone na tela, correspondendo, por exemplo, à consciência. Clicamos o ícone e nosso computador adquire mente e consciência[1].

Nunca se chegou a um consenso acerca de como e até que ponto gostaríamos de povoar nosso universo, com mentes e com cérebros ou apenas com cérebros. Nessa disputa interminável, a estratégia do dualista foi sempre a de tentar encontrar uma marca distintiva do mental, algo que nos levasse a acreditar que ele não apenas é diferente, como também irreconciliável com o físico ou com o sensível e observável. Seu desejo é o de encontrar uma assimetria fundamental entre mente e cérebro, uma assimetria da qual se pudesse derivar a plausibilidade de crenças religiosas na realidade e imortalidade das mentes ou almas. Disto decorre uma adoção implícita, mas inevitável, de algum tipo de platonismo por parte do dualista – um platonismo que serviria para fundamentar uma assimetria entre mente e cérebro, como, por exemplo, o caráter eterno, indestrutível e imutável que seria comum às mentes e às verdades matemáticas. A proposta desse tipo de assimetria não parece, entretanto, ter se revelado tão boa assim: a física mostra que a matéria é tão indestrutível como o são as mentes. E a história da ciência mostra que as verdades matemáticas não são tão eternas e imutáveis quanto poderíamos supor. Em geometrias não euclidianas há triângulos cuja soma dos ângulos não têm 180 graus.

O monista caminha na direção contrária, procurando desfazer as assimetrias e assimilar o mental ao físico. Sua proposta e sua motivação coincidem com a ambição científica que caracteriza nosso século. Apostamos no triunfo de uma ciência que dê conta de todos os fenômenos que nos rodeiam; uma ciência que unifique a diversidade do mundo num único padrão explicativo. Trata-se então de explicar *cientificamente* a natureza do mental, ou, em outras palavras, explicar o mental em termos de fenômenos físicos ou cerebrais. Essas tentativas, contudo, esbarram em grandes dificuldades e acabam revelando a existência de muito mais assimetrias entre men-

1. Este tipo de fantasia é descrito, de forma brilhante, por Alexander (1996).

te e cérebro do que se imaginava inicialmente – assimetrias que ainda não podem ser explicadas pela ciência.

Consideremos novamente o neurocientista examinando o cérebro para tentar encontrar nesse algo que se assemelhe a estados mentais. Hoje em dia dispomos de técnicas mais aperfeiçoadas para realizar esse tipo de investigação. Contamos com vários recursos para produzir imagens do cérebro em funcionamento ou mesmo para medir sua atividade elétrica. Um deles é o eletroencefalograma ou EEG. Através do EEG podemos determinar, durante o sono, quando alguém está sonhando, ou seja, quando alguém entra em sono REM, pois, nesse caso, o eletroencefalograma se altera substancialmente[2].

A dificuldade surge na medida em que, pelo exame do EEG, podemos saber que o indivíduo está sonhando, mas não podemos saber *com o que* ele está sonhando. Da mesma maneira, podemos, pelo exame da atividade química e glandular do corpo de uma pessoa, saber se ela está tendo um ataque de fúria. Haverá mais adrenalina no sangue dessa pessoa. Mas, da mesma forma, a detecção de uma maior quantidade de adrenalina no corpo de uma pessoa permite-nos saber que ela está tendo um ataque de fúria, mas não saber com *o que* ou *com quem* ela está enfurecida. Ora, a questão que se coloca é a seguinte: será o exame da atividade física do corpo ou do cérebro suficiente para determinar os conteúdos mentais que ocorrem a essa pessoa? Ou haverá um hiato intransponível entre cérebro e estados subjetivos, um hiato que se impõe pela incapacidade de se estabelecer um caminho entre os sinais elétricos do cérebro, sua atividade química e aquilo que podemos identificar como sendo nossos conteúdos mentais ou nossas ideias?

Não dispomos de respostas para essas questões. A ciência ainda não conseguiu resgatar esse hiato, que se alarga ainda mais quando supomos que esses eventos mentais são *experiências conscientes*. O problema que se coloca é, então, não apenas o de traçar algum tipo de correlação entre duas séries (a do físico e a do mental), mas o de saber como a série de eventos físicos pode produzir o aspecto específico desse tipo de experiência que a torna uma *experiência consciente*. Em outras palavras: se há uma relação entre esses dois tipos de séries, que tipo de relação será essa? Como passamos de um conjunto de propriedades para outro, aparentemente tão distinto? O que falta é algum tipo de explicação que torna *inteligível* a passagem entre o físico e o mental – uma inteligibilidade que requer mais do que simplesmente estabelecer correlações. Que tipo de propriedade ou que tipo de circunstância leva a matéria (o cérebro) a produzir consciência?

Responder a essa questão parece ser o grande desafio a ser enfrentado por aqueles que apostam no monismo ou na possibilidade de que a ciência possa fornecer uma explicação definitiva da natureza da mente e de suas relações com o cérebro. Trata-se de

2. REM são as iniciais de Rapid Eye Movement, um estágio no sono onde ocorre movimento ocular acompanhado de sonhos vívidos. O REM produz modificações detectáveis no EEG.

explicar como o cérebro pode produzir fenômenos mentais subjetivos ou conscientes utilizando-se das mesmas categorias explicativas que são usadas para explicar o funcionamento de um sistema físico. Que tipo de *interpretação física* podem ter fenômenos como a subjetividade e a consciência?[3] Serão as leis naturais suficientes para capturar e explicar os aspectos específicos que levam à produção da subjetividade e da consciência?

Tomemos mais um exemplo. Suponhamos novamente que um neurocientista observa o cérebro de alguém e que ele queira *identificar* os estados mentais dessa pessoa com a atividade exibida pelo seu cérebro, da mesma maneira que um químico identifica água com H_2O. Esse tipo de identificação é frequentemente buscada pela investigação científica pois supõe-se que através dela se possa chegar a algum tipo de explicação da natureza de um determinado fenômeno – por exemplo, uma explicação do *que é* a água. O neurocientista estaria procurando uma explicação do que são estados mentais através de uma identificação desses com estados cerebrais – uma identificação entre os dois modos de apresentação ou as duas descrições possíveis de um mesmo fenômeno. Digamos que, após algumas investigações, ele conclua que o estado mental correspondente a sentir que uma determinada dor ocorre sempre que as fibras-C do sistema nervoso forem estimuladas. Ora, será que podemos afirmar que "estimular as fibras-C" significa *explicar* o que é sentir uma determinada dor? Até que ponto a descrição de uma dor como "estimulação das fibras-C" realmente expressa os aspectos subjetivos e conscientes envolvidos em sentir uma determinada dor? Ou, em outras palavras, será que a descrição "estimular as fibras-C" poderia expressar o que significa sentir uma determinada dor?

O obstáculo que o neurocientista enfrenta é a transposição de uma descrição *em terceira pessoa* (estimulação das fibras-C) para uma descrição *em primeira pessoa*, onde se expressam estados subjetivos e experiências conscientes. Esse é o problema da interpretação física de fenômenos que envolvem subjetividade e consciência – um problema que sugere que esses fenômenos nunca poderiam ser integralmente descritos ou reduzidos aos termos de uma linguagem científica. A *relação* entre o mental e o físico não seria capturada por uma linguagem construída exclusivamente em terceira pessoa e descrevendo unicamente eventos públicos, como é o caso da linguagem da ciência – uma linguagem onde a conexão entre estados mentais e estados cerebrais não nos permite aproximá-los ao ponto de podermos dizer, por exemplo, que sentir uma determinada dor e estimular as fibras-C são a mesma coisa, da mesma maneira que água e H_2O o são.

Entender essa relação e poder expressá-la na linguagem da ciência seria uma tarefa importantíssima – uma tarefa que ultrapassa, em importância, a pura e simples escolha de uma imagem do mundo por contraposição a outra. Precisaríamos saber não ape-

---

3. Este termo é introduzido por McGinn (1982). Neste capítulo seguimos, em grande parte, a apresentação do problema mente-cérebro sugerida por esse autor.

nas como de cérebros podem emergir mentes, mas como essas últimas podem, reciprocamente, alterar o cérebro e o corpo. Sabemos que a ordem física dos eventos cerebrais altera estados mentais. Como dissemos há pouco, esse é um conhecimento intuitivo, que temos quando tomamos bebidas alcoólicas e drogas que alteram o equilíbrio químico das reações entre as diversas partes do cérebro. Mas não sabemos como, a partir de sinais elétricos, passamos aos pensamentos. Não sabemos tampouco como esses podem alterar os próprios sinais elétricos do cérebro e influenciar nosso corpo a ponto até de gerarmos vários tipos de doenças ou disfunções orgânicas. Novamente o problema que se coloca é nossa incapacidade não apenas de descobrir, mas talvez de *conceber* como se dá a passagem entre o mental e o cerebral, entre o físico e o subjetivo.

Além daqueles que apostam na ciência e no triunfo futuro do monismo e daqueles que apostam no dualismo, seja abraçando algum tipo de crença religiosa ou não, podemos identificar um terceiro grupo nesse debate: os que supõem que o problema da relação mente-cérebro não pode ser resolvido; os chamados "agnósticos". Assim como na matemática há vários problemas cuja solução é impossível, o mesmo ocorreria com o problema das relações entre mente e cérebro. Os agnósticos partem da ideia de que o problema da interpretação física dos fenômenos subjetivos e conscientes seria agravado pelo fato de nossa vida mental não nos permitir acesso senão aos *conteúdos mentais* que a compõem, ou seja, que nosso pensamento nunca poderia nos fornecer alguma pista acerca dos processos cerebrais que estão envolvidos na sua própria produção. Seria pouco provável que algum dia essa condição pudesse ser alterada: a perspectiva que podemos ter do mundo nos confina ao nosso próprio universo mental.

Para o agnóstico, uma imagem do meu próprio cérebro em funcionamento – como aquelas obtidas por ressonância magnética altamente sofisticada ou por qualquer outra técnica de neuroimagem – apresenta-se tão estranha a mim mesmo quanto uma radiografia de meu pulmão. Eu só saberia que aquela é uma imagem do *meu* próprio cérebro em funcionamento ou que aquela é uma radiografia de *meu* pulmão se alguém me fornecesse essa informação. Eventos neurais, ou seja, a interpretação física de fenômenos mentais não participa de nossas experiências subjetivas; podemos no máximo traçar algumas correlações, mas essas não explicam a *passagem* entre o físico e o subjetivo.

Imagine uma situação experimental onde eu sou convidado junto com várias outras pessoas a pensar sobre uma mesma coisa, ou seja, produzir conteúdos mentais semelhantes. Imagine também que, simultaneamente, alguém esteja produzindo imagens desses vários cérebros em funcionamento – imagens que, por hipótese, mostrariam em todos esses cérebros um mesmo conjunto de áreas sendo estimuladas. Todos veríamos simultaneamente essas imagens projetadas numa tela diante de nós – sem saber, contudo, qual imagem corresponde ao cérebro de quem, ou seja, as imagens não seriam numeradas nem posicionadas de forma a dar esse tipo de pista. E todas as imagens seriam praticamente iguais, mostrando um cérebro com algumas regiões semelhantes em atividade. Seria eu capaz de dizer qual dessas imagens corresponde à do *meu* cérebro, ou seja, seria eu capaz de relacionar meu pensamento com a imagem de minha atividade cerebral a não ser que alguém me informasse que aquela era a imagem

do *meu* cérebro? A imagem do meu cérebro é tão estranha quanto a imagem do meu fígado; só sei que aquela é a imagem de meu fígado se alguém me disser isso ou se eu estiver numa máquina de ultrassom sabendo, de antemão, que as imagens que aparecem na tela são do meu fígado. Esse experimento com imagens de vários cérebros projetadas simultaneamente numa mesma tela – que, talvez por razões práticas, nunca pudesse vir a ser realizado – ilustraria a pressuposição do agnóstico, qual seja, a de que jamais poderíamos encontrar uma passagem entre sinal cerebral e experiência consciente. Seria mais uma versão daquilo que os filósofos chamaram de problema da intransponibilidade entre primeira e terceira pessoa e que expressaria toda a dificuldade envolvida no problema da relação entre mente e cérebro.

O desânimo do agnóstico implicitamente nos convida a abandonar o monismo e optar pelo dualismo. Essa é uma visão radicalmente oposta àquela sugerida pela neurociência. Para o dualista, a mente é imaterial, com propriedades distintas e incompatíveis com o mundo físico. Será que podemos sequer imaginar o que seria algo imaterial? Certamente só poderíamos conferir uma caracterização negativa para esse conceito, ou seja, só poderíamos definir algo imaterial por oposição às propriedades da matéria, ou seja, dizendo o que essa substância imaterial *não é*. Dizendo que ela não tem localização espacial, não tem peso, não tem massa etc. Eis aqui um belo exercício filosófico para desafiar nossa imaginação.

Aparentemente, o dualista não precisaria se preocupar com o problema da interpretação física dos fenômenos subjetivos e conscientes. Pois ele não busca saber como poderíamos caracterizar esse tipo de fenômenos a partir das descrições em terceira pessoa, utilizadas pela linguagem científica: seu ponto de partida é considerar subjetividade e consciência como distintos e irredutíveis a qualquer tipo de base física. Contudo, o problema da relação entre o mental e o físico persiste, embora de uma maneira diferente. A maior dificuldade enfrentada pelo dualismo é saber como algum tipo de relação entre mente e cérebro pode ocorrer: como conceber que algo imaterial possa, de alguma maneira, afetar coisas materiais como o cérebro? Nesse sentido, a única diferença entre monismo e dualismo seria uma inversão na direção da investigação. O primeiro teria de dar conta de como a partir do mundo material pode surgir subjetividade e consciência, ao segundo caberia mostrar como subjetividade e consciência poderiam afetar o mundo material.

Esta última questão não é menos importante e nem fácil de responder, sobretudo quando se considera o papel de intenções, crenças e desejos – ou seja, estados mentais – na sua relação com o comportamento. Se esses estados mentais são imateriais, como poderiam afetar o curso de nossos comportamentos? Não parece intuitivo que nossos comportamentos sejam, de alguma forma, causados ou guiados por nossas intenções ou desejos? Mas, nesse caso, como algo imaterial pode afetar nosso corpo? Como minha intenção ou meu desejo de levantar e ir abrir a porta poderia causar algo no mundo físico como, por exemplo, meus movimentos musculares? Ou teremos que acreditar que nossos comportamentos não são causados por nossas intenções e desejos? Teríamos então de abrir mão da noção intuitiva de que nossos comportamentos resultam de

estados da minha mente e com ela mantém algum tipo de relação causal? Ou buscar uma outra maneira de conceber a própria relação de causalidade? Não seria mais prudente abandonar o dualismo?

Há ainda outras dificuldades que precisam ser enfrentadas pelo dualista: se a mente é imaterial e totalmente independente do cérebro, como explicar que danos causados a este último possam também afetar atividades mentais? E, se a mente é imaterial e independente do cérebro, por que temos então um cérebro tão complexo comparado com o de outros seres vivos?

O dualismo é visto pelos filósofos da mente contemporâneos como uma doutrina metafísica extravagante – uma doutrina que dificilmente poderia ser coerentemente sustentada. Mas talvez o aspecto mais problemático do dualismo seja o fato de ele ser uma filosofia *sem agenda*. Ou seja, tudo que o dualista pode fazer é tentar provar a existência de uma diferença radical entre mente e matéria. Daí para diante nada mais poderia ser feito. Como afirmamos acima, a ideia de algo imaterial só pode ser caracterizada negativamente, por oposição às propriedades do mundo material: nada poderia ser afirmado acerca da natureza do mental além do fato de ele ser distinto do físico. Nunca uma ciência do mental poderia ser desenvolvida, pois a mente não teria nenhuma característica que permitisse qualquer tipo de abordagem científica.

Ora, haverá alternativas ao monismo e ao dualismo sem que essas signifiquem desistir de tomar uma posição como faz o agnóstico? Ou serão essas as duas únicas alternativas possíveis para tentarmos resolver o problema mente-cérebro? Antes de tentarmos responder essas duas questões precisamos caracterizar o *tipo* de problema com que nos defrontamos quando consideramos a relação entre mente e cérebro e saber por que ele tem persistido ao longo da história da filosofia e da história da ciência.

Acreditamos ser organismos complexos e que nossas funções mentais se devem a essa complexidade. Ou seja, somos educados de acordo com uma tradição científica que sempre esteve presente na nossa cultura. Ao mesmo tempo recebemos uma educação religiosa, que nos ensina sermos dotados de espíritos que sobreviverão após nossa morte. Como se não bastassem essas contradições, somos ainda educados a usar dois tipos de vocabulário: um vocabulário físico e um vocabulário mental. Referimo-nos a eventos físicos usando um tipo de vocabulário; para os eventos mentais desenvolvemos um vocabulário específico, que sugere, implicitamente, que o físico e o mental são distintos. Só esses dois fatores já seriam suficientes para percebermos por que a relação entre mente e cérebro teria, necessariamente, de se apresentar como um problema para nós[4].

Defrontamo-nos, na verdade, com duas crenças contraditórias, mas nenhuma delas pode ser considerada ingênua. Por um lado somos levados a crer no monismo e na

4. Encontramos a mesma observação em Hannan (1994). Os parágrafos abaixo seguem o percurso dessa autora que também ressalta que o problema mente-cérebro é essencialmente conceitual.

aposta de que o problema mente-cérebro é um problema científico, ou seja, um problema empírico que poderia ser resolvido, algum dia, através de alguma descoberta científica – da mesma maneira que se descobriu, por exemplo, a cura das infecções através da penicilina. A característica básica dos problemas científicos consiste no fato de que eles podem ser resolvidos através da observação e da experimentação. Ora, conforme sugerimos, essa estratégia pode rapidamente se defrontar com várias limitações. É possível que, por mais que a neurociência avance, isto é, por mais que ela nos proporcione dados empíricos ou resultados experimentais, esses se limitem a ser sempre informações apenas acerca do funcionamento do cérebro. O problema que pode permanecer – e do qual já falamos aqui ao discutir a questão da interpretação física dos fenômenos subjetivos e conscientes – seria saber como poderíamos relacionar todos esses dados e observações com os aspectos subjetivos presentes na nossa vida mental. É possível que nenhum dado *em si* possa nos ajudar a passar de uma perspectiva de terceira pessoa para uma de primeira pessoa. Mesmo que esse dado venha a surgir, sempre poderá existir algum tipo de discussão envolvendo sua interpretação. Ao incluirmos a necessidade de uma interpretação específica já nos afastamos da proposta de que algum tipo de dado poderia, por si só, resolver o problema da relação mente-cérebro. Contudo, o problema da interpretação e o problema da passagem da terceira para a primeira pessoa não parecem ter sido suficientes para abandonarmos nossa aposta no triunfo do monismo e na explicação neurocientífica da natureza dos fenômenos mentais. Prova disto é o fato de que a pesquisa neurocientífica continua se desenvolvendo cada vez mais na sua busca pelos correlatos neurais da consciência e da experiência subjetiva.

Adotar uma perspectiva oposta ao monismo significaria acreditar no que dizem os dualistas e os agnósticos. Vimos que essas duas posições, embora distintas, tendem a se confundir, pois nossa tendência é interpretar o discurso do agnóstico – um discurso pela impossibilidade de solução do problema mente-cérebro – como implicando na verdade o do dualismo. Desistir de resolver um problema, ou concluir que a ciência não pode resolver o problema mente-cérebro, não implica, porém, em optar pela solução contrária. É possível ser agnóstico sem ser dualista. Quando um matemático conclui que um determinado problema não tem solução ou que sua solução é impossível ele não está necessariamente sugerindo que devemos jogar fora o conhecimento matemático. O difícil, porém, é ser dualista sem abraçar, simultaneamente algum tipo de concepção religiosa: embora uma coisa não implique necessariamente na outra, nossa tradição cultural sugere esse tipo de aliança.

O monista conta com dados, observações e métodos experimentais consagrados. O dualista conta apenas com argumentos filosóficos e crenças religiosas. Vista dessa perspectiva, nossa escolha por uma doutrina em favor de outra já estaria feita. Restaria apenas aguardar que a ciência se desenvolvesse e culminasse com uma solução definitiva para o problema mente-cérebro. O monista não conta com a possibilidade de que isto possa *não acontecer* ou que a própria ciência possa concluir que a solução desse problema seja impossível. O monismo tem sido uma grande aposta no futuro da ciência – uma aposta que, aliás, já dura alguns séculos. O monista acredita no sucesso futu-

ro da ciência da mesma maneira que os dualistas (na sua maioria) acreditam na religião: ambos fazem discursos opostos, nos quais, entretanto, se expressa algum tipo de crença. Faz parte da crença do monista acreditar que não existem problemas para os quais a ciência não possa oferecer uma solução. Acreditar na existência desse tipo de problema não significa, aliás, como sugere o monista, assumir necessariamente uma postura anticientífica.

Não podemos esperar que a ciência responda, por exemplo, questões do seguinte tipo: "Quais os princípios que devem reger uma sociedade para que essa seja justa?" Tal questão não poderia ser respondida unicamente com base em dados de observação ou em experimentos. Isto não a torna uma falsa questão nem tampouco uma questão irrelevante. Trata-se de uma questão de outra natureza, ou seja, aquilo que chamamos de uma *questão conceitual*. Questões conceituais dizem respeito ao modo como concebemos o mundo e como podemos tornar essas concepções coerentes e adequadas. Reconhecer a existência de questões conceituais não significa desprezar a investigação científica nem os dados que podem se originar dessa. Significa apenas reconhecer que dados e experimentos podem não ser suficientes para resolver certas categorias de problemas.

A ilusão do monista consiste em achar que a ciência poderia resolver todos os problemas, inclusive os conceituais. Ao assumir essa perspectiva, ele se esquece de que ao fazer ciência ele está, necessariamente, envolvendo-se com questões conceituais. É ingênuo supor que podemos separar a produção da ciência de questões conceituais, embora muitas vezes o ensino de ciências que recebemos na escola nos dê a impressão de que isto seria possível. Questões conceituais perpassam a própria prática da ciência e compõem os pressupostos sobre os quais essa se baseia. Basta pensar nos *conceitos* que são implicitamente mobilizados pelo discurso científico, como, por exemplo, *teoria*, *explicação*, *confirmação*, *teste empírico* e assim por diante.

Usar a ciência para explicar a natureza dos conceitos sobre os quais ela mesma se assenta equivaleria a correr o risco de incorrer num círculo vicioso: dificilmente chegaríamos a algum lugar. O mesmo poderíamos afirmar das tentativas de resolver o problema das relações entre mente e cérebro na qualidade de um problema unicamente científico ou empírico: pois a mente que se quer explicar é a mesma que produz a ciência usada para explicá-la. Estaríamos girando em círculos.

Os dualistas e os agnósticos correm, entretanto, o risco inverso. Eles supõem ser possível discutir e até chegar a uma solução para o problema da relação entre mente e cérebro sem sair de sua poltrona, isto é, virando as costas para a ciência e para qualquer tipo de resultado empírico que possa surgir dessa. Eles parecem se esquecer de que a filosofia começa onde a ciência acaba, ou seja, que os problemas conceituais adquirem sentido e consistência por se situarem no limiar da investigação científica. Isto é o que vem ocorrendo, por exemplo, com a física no século XX que passou a se confrontar com questões do tipo "será o universo determinista ou não?" ou "podemos ter uma imagem estritamente objetiva do mundo físico?" e assim por diante. Questões do tipo

"qual a natureza da mente?" ou "como é possível a relação entre mente e cérebro?" surgem no limiar da neurociência. Avaliar sua verdadeira amplitude e espessura não poderia ser feito se ignorássemos o que a ciência tem a dizer acerca da mente, do cérebro e de sua relação.

O intercâmbio entre a análise conceitual e a investigação científica configura, assim, a trilha a ser seguida pela filosofia da mente. Resultados científicos por si só não esgotam a verdadeira dimensão teórica e filosófica envolvida no problema da relação entre mente e cérebro. Mas não podemos prescindir desses se quisermos formular esse problema com a precisão necessária para desenvolver seriamente a análise conceitual.

O primeiro passo nesse percurso consiste em mapear os modos possíveis de conceber as relações entre mente e cérebro que a filosofia da mente pode nos proporcionar. Eles são apresentados no quadro abaixo:

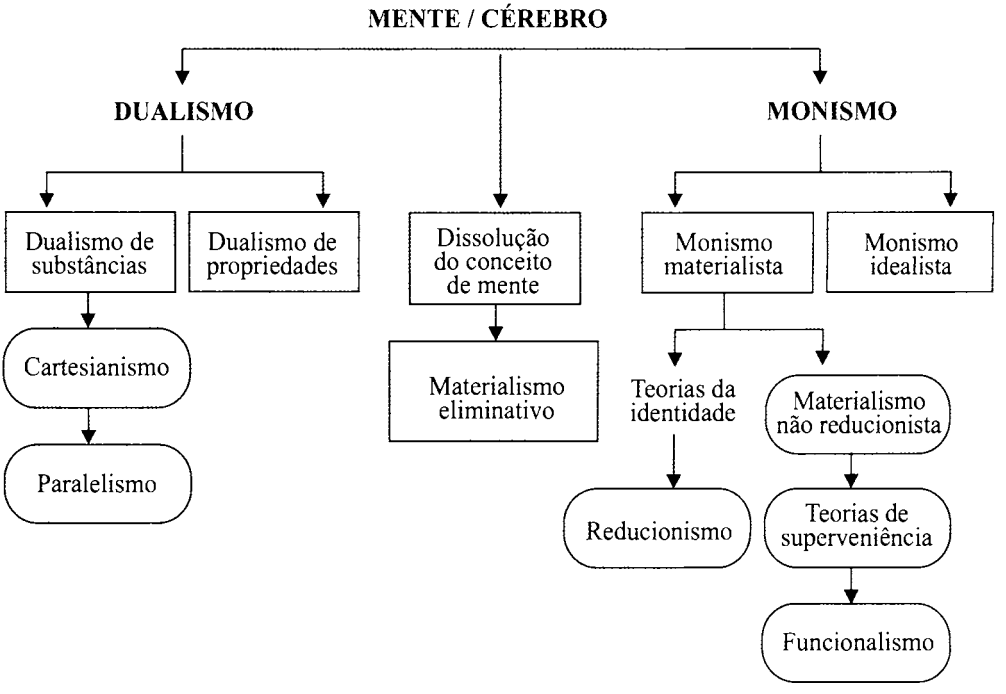

Fig. 1.2 – Esquema geral das relações entre mente e cérebro.

As ramificações à esquerda representam os principais tipos de dualismo. As da direita, as variedades do materialismo, e as do centro as tentativas de resolver o problema mente-cérebro tratando-o como um pseudoproblema que se dissiparia pela dissolução do próprio conceito de mente. Antes de começar a percorrê-las abordaremos o dualismo de substâncias numa versão específica, qual seja, o cartesianismo. Começar pelo cartesianismo significa não apenas revisitar uma filosofia do século XVII, mas

buscar as raízes históricas do problema mente-cérebro, refazendo um itinerário teórico de onde se derivaram concepções ainda presentes no horizonte da neurociência e da filosofia da mente contemporâneas.

O QUE LER

CHURCHLAND, P. *Matter and Consciousness*

HEIL, J. *Philosophy of Mind: a contemporary introduction*

MCGINN, C. *The Character of Mind*

CAPÍTULO II

A HERANÇA CARTESIANA

Foi René Descartes (1596-1650) quem propôs, pela primeira vez nos tempos modernos, a existência de uma descontinuidade essencial entre mente e corpo, entre o físico e o mental. A partir de sua filosofia (o *cartesianismo*), a questão da separação entre matéria e pensamento torna-se um *problema filosófico*. O cartesianismo formula e institui esse problema específico, a partir de uma demonstração filosófica na qual Descartes supõe que podemos *deduzir*, numa cadeia de raciocínios coerentes, que corpo e alma são duas substâncias distintas, e que suas propriedades são incompatíveis.

Descartes estabelece essa cadeia de raciocínios dedutivos a partir do *Cogito*. Embora nunca tenha dito o que é pensar e muito menos o que é existir, Descartes toma como certeza primeira, basilar, a proposição *Penso, logo existo*. Isto porque podemos duvidar de qualquer coisa, até mesmo se o mundo existe ou se 2 + 2 = 4, mas não podemos *duvidar que duvidamos*, ou seja, não podemos duvidar que *pensamos* ao formular nossas próprias dúvidas, pois dúvidas são pensamentos. Assim sendo, é impossível *pensar que não pensamos*, pois nesse caso estaríamos incorrendo numa contradição. *Penso, logo existo* é uma proposição única e peculiar, na medida em que não é possível negá-la.

Descartes conseguiu demonstrar, assim, que a existência do pensamento e do espírito que o pensa constitui a única e verdadeira certeza de que dispomos se duvidarmos de tudo. Dessa proposição, "Penso, logo existo" – cuja verdade teríamos de aceitar sob quaisquer circunstâncias – ele deriva a ideia de que espírito e corpo são radicalmente distintos, formulando dois tipos de argumentos:

1) A mente deve ser diferente do corpo na medida em que essa é mais fácil de ser conhecida;

2) Substâncias materiais são divisíveis; o mesmo não se aplica ao mental.

O primeiro argumento baseia-se na pressuposição de que o espírito é mais fácil de conhecer do que o corpo. Se podemos conhecer uma coisa melhor do que outra, isto

significa que deve existir uma grande diferença entre essas duas coisas. Logo – argumenta Descartes – espírito e corpo devem ser diferentes. Como veremos mais adiante, esse argumento antecipa a chamada doutrina da acessibilidade direta, proposta na filosofia da mente contemporânea.

O segundo argumento estipula que o físico e o mental são assimétricos, pois qualquer coisa extensa no espaço (como é o caso dos corpos) é infinitamente divisível. Ora, as ideias que estão no meu espírito não são divisíveis como um corpo físico. Não teria cabimento dizer que poderíamos cortar um pensamento da mesma maneira que cortamos uma fatia de pão. Ao fazer essa afirmação, Descartes atribui ao mental uma característica radicalmente diferente daquela possuída pelos corpos físicos: a não espacialidade. Pensamentos não são *coisas extensas* que ocorreriam no espaço; é nesse sentido que eles são indivisíveis por oposição a qualquer porção de matéria e que não seria possível *localizar* pensamentos nessa ou naquela parte do espaço, nem sequer no nosso próprio corpo. Indivisibilidade e não espacialidade são propriedades do mental que o tornam radicalmente assimétrico em relação ao físico. Se isto não fosse verdade, quando amputo uma perna eu deveria estar, ao mesmo tempo, amputando uma parte de minha alma. Certamente Descartes nunca imaginou nem presenciou o que aconteceria à nossa atividade mental se amputássemos parte do cérebro. No século XVII não havia nada parecido com a neurociência de que dispomos hoje.

A questão da separação entre matéria e pensamento torna-se particularmente *problemática* a partir da obra de Descartes, na medida em que ela envolve saber como seria possível a relação entre uma alma imaterial e um corpo físico, e como ambos poderiam se influenciar apesar de serem radicalmente diferentes. Ora, se espírito e corpo são radicalmente diferentes, isto é, se o espírito é imaterial, como ele poderia interagir *causalmente* com o corpo? Em outras palavras, como é que a intenção de levantar-me e abrir a porta, que é imaterial como tudo que ocorre no meu espírito, poderia interagir (causalmente) com meu corpo, produzindo um movimento nos meus músculos que me poria a andar e caminhar até a porta?

Havia um pressuposto bastante forte em toda a filosofia cartesiana: uma fé inquestionável na veracidade do princípio da causa e efeito, ou, em outras palavras, no *princípio de causalidade*. Sem o princípio de causalidade, Descartes não poderia sustentar sua visão de um universo mecânico onde tudo funcionaria através de causa e efeito – um princípio fundamental que deve nortear nossos raciocínios e investigações científicas. Inspirado na física do século XVII, Descartes acreditava num universo mecânico. Para se ter uma boa imagem do que é universo mecânico e o papel fundamental que nele tem o princípio de causalidade, basta imaginarmos um dispositivo do tipo de uma alavanca e engrenagem. Cada vez que se puxa a alavanca, isto causa um movimento na engrenagem. O universo seria um imenso sistema mecânico – algo como uma imensa relojoaria criada por Deus – onde tudo seria governado pela lei de causa e efeito. Tudo no universo funcionaria dessa maneira, inclusive nós mesmos. A impossibilidade de imaginar algum tipo de interação causal entre mente e corpo comprometeria o caráter universal do princípio de causalidade. Foi isto que fez com que a relação mente-corpo se tornasse um *problema*.

Uma incursão pelas "Meditações" de Descartes

Para compreendermos como Descartes chegou a formular esses dois argumentos a favor de uma distinção radical entre matéria e pensamento precisamos refazer, ainda que de forma sucinta, o percurso seguido nas suas *Meditações*, publicadas em 1641. Nesse trabalho – que mais parece um drama filosófico para o leitor desavisado –, composto de seis capítulos ou seis "meditações", Descartes expõe as razões que o levaram a sustentar que a alma é distinta do corpo ou por que pensamento e matéria seriam incompatíveis.

Nenhum filósofo apostou e, ao mesmo tempo, explorou tanto os aspectos paradoxais que estão envolvidos na nossa experiência introspectiva. Da experiência introspectiva chega-se a um paradoxo ou a uma contradição, da contradição salta-se para uma certeza autoevidente ou uma primeira certeza de onde, através de uma cadeia de raciocínios dedutivos, pode-se ir derivando, passo a passo, outras certezas. Esse é o sonho do edifício racional perfeito, inabalável. E a crença absoluta nos poderes da razão como instrumento para se chegar a um conhecimento seguro. Para se obter esse conhecimento seguro reverte-se a ordem habitual de começar pela experiência das coisas para depois raciocinar sobre elas. Ao contrário, deve-se começar pela razão para ver o que essa pode ensinar e depois dela derivar um conhecimento sobre o mundo. Essa é uma reversão completa dos nossos métodos habituais de se obter conhecimento, uma reversão *metodológica* de consequências profundas.

Partir da razão para saber o que ela pode nos ensinar significa, num primeiro momento, exercitar um de seus maiores poderes, qual seja, a possibilidade de colocar em dúvida tudo o que se sabe até então. A dúvida é a primeira grande expressão do poder da razão. Pode-se duvidar de tudo, a começar daquilo que nos é transmitido pelos sentidos, por nossas sensações, questionando até que ponto essas seriam confiáveis. A dúvida vai demolindo as certezas habituais, num processo progressivo. Posso duvidar das minhas sensações, duvidar até mesmo se o mundo que me é dado pelas sensações seria uma realidade ou apenas uma fantasia da minha mente. A dúvida sistemática ou a dúvida hiperbólica, como diria Descartes, seria o instrumento da minha razão para combater um gênio maligno, uma figura alegórica que simbolizaria a tentativa persistente e habilidosa de meus sentidos e de meus próprios raciocínios, que poderiam levar-me ao engano ou ao falso conhecimento. Esse é o grande tema do primeiro capítulo das *Meditações* ou a primeira meditação de Descartes.

Na primeira meditação encontramos o famoso argumento do sonho: minhas sensações, quando estou acordado, são tão vívidas como aquelas que tenho quando estou sonhando. Assim sendo, como poderia eu distinguir entre sonho e vigília? Não haveria nenhuma marca, do ponto de vista de minha experiência interna, que me permitiria saber se as impressões que tenho estão ocorrendo durante minha vigília ou durante um sonho. Nada me garante que eu esteja acordado quando penso estar; meu sonho teria o poder de me convencer até mesmo de que eu estaria acordado quando sonho. Ou seja, quando penso que estou acordado, poderia estar sonhando um sonho no qual tudo se passaria como se eu estivesse acordado. Minha experiência interna, introspectiva, é

insuficiente para me fornecer algo a partir do qual eu pudesse decidir se estou acordado ou estou sonhando.

Se a primeira meditação tivesse de ser escrita hoje em dia, possivelmente a ideia de um gênio maligno seria representada na forma de um neurocientista perverso que, através de um implante cuidadoso de eletrodos no meu cérebro, poderia produzir em mim vários tipos de sensações, a começar por sensações visuais ou experiências de estar percebendo alguma coisa diante de mim – mesmo que eu estivesse momentaneamente cego. Eu perceberia coisas, a despeito de elas não estarem diante de mim e eu não poder enxergá-las. Esse neurocientista tornar-se-ia um verdadeiro gênio maligno se, através do implante desses eletrodos, passasse a produzir em mim, não apenas experiências visuais de caráter episódico como também reproduzisse o modo como essas são encadeadas numa sequência que imitasse perfeitamente uma percepção real e ordenada do mundo. Certamente esse neurocientista teria de desenvolver uma técnica bastante sofisticada, para saber exatamente o local e a sequência em que os eletrodos teriam de ser implantados em meu cérebro, de modo a produzir uma alucinação tão bem estruturada que eu nunca poderia saber se estaria alucinando ou não. Ter alucinações indiscerníveis em relação a experiências visuais reais levar-me-ia a formular o argumento do sonho. Esse me colocaria numa situação de dúvida atroz, mas não seria propriamente um problema para o neurocientista chegar a estabelecer uma confusão entre sonho e realidade, entre percepção real e onírica. Por uma estranha coincidência – ou talvez um fato revelador do verdadeiro toque de gênio que existe na obra de Descartes – estudos neurofisiológicos recentes sugerem que a atividade cerebral subjacente à produção de experiências visuais sonhadas e experiências que temos durante a vigília é exatamente a mesma! A fantasia de Descartes, seu gênio maligno, ainda seria real, a despeito de suas *Meditações* terem sido publicadas há mais de 300 anos![1]

O passo seguinte, que inaugura a segunda meditação, consiste em mostrar como a dúvida introspectiva leva-nos inevitavelmente a um paradoxo ou uma autocontradição – uma autocontradição que nos permite, entretanto, saltar para a primeira certeza. A dúvida, levada ao extremo, devora sua própria cauda. Não posso duvidar que duvido; disto tira Descartes sua frase mais célebre, o "Penso, logo existo" ou o "*Cogito ergo sum*". Se há um neurocientista que me produz alucinações, tudo pode ser uma alucinação, exceto esse próprio neurocientista. Ele não pode ser uma alucinação e produzir uma alucinação. Deve haver algo fora de mim, para além de meus estados subjetivos. Descartes não pensava em termos de um neurocientista produzindo alucinações, mas de um "eu" que não poderia *produzir* alucinações se ele mesmo não passasse de uma alucinação. Esse é o passo mais importante na sequência das *Meditações*, mas também o mais delicado. Pois do paradoxo de não poder duvidar sobre a dúvida, Descartes derivou a existência de um "eu" real, de uma *substância pensante*, que, embora diferente da matéria, seria algo como

1. Cf. LLINÁS, R.R. & PARE, D. (1991).

uma substância, ou seja, uma *substância* imaterial. Mais de 300 anos depois, outro filósofo questionaria se os paradoxos da dúvida teriam provado efetivamente a existência de uma substância (mesmo que imaterial) ou apenas a certeza de existir um pensamento acerca dessa substância[2].

A terceira meditação é normalmente interpretada como o esforço de Descartes para provar, a partir do *Cogito*, a existência de Deus. Mas ela é, na verdade, uma grande reflexão sobre a natureza do infinito, ou, em outras palavras, uma tentativa de explicar como que de uma mente finita e submetida à finitude de suas próprias experiências que, por sua vez, só descortinam um mundo finito, poder-se-ia formar uma *ideia de infinito*. Descartes afirma que somente Deus poderia ter colocado em nossa mente a ideia de infinito e de que o infinito, portanto, existe. Esse, entretanto, é ainda um problema bastante atual para os cientistas cognitivos contemporâneos: saber como que um cérebro finito pode gerar algo infinito, como, por exemplo, uma sequência infinita de sentenças na nossa linguagem. A ciência cognitiva contemporânea chama isto de *recursão*, ou o "truque que multiplica pensamentos humanos em números verdadeiramente astronômicos"[3].

Mas é somente na sexta meditação que Descartes proporá, mais explicitamente, seu dualismo metafísico, isto é, sua concepção de que o corpo é radicalmente distinto da alma. O mental é indivisível, o mesmo não ocorre com o material. Qualquer segmento da matéria, na medida em que é extenso, é também divisível: encontramos aqui uma assimetria ou uma incompatibilidade entre as propriedades que caracterizam pensamento e matéria.

A sexta meditação já coloca o problema de como relacionar o mental (imaterial) com o físico (o corpo extenso). No livro *Les Passions de l'âme* [As paixões da alma], publicado em 1649[4], Descartes explicita a proposta da glândula pineal como sede da alma. A glândula pineal seria uma espécie de interface entre mente e corpo: os movimentos físicos do sistema corporal moveriam a glândula que, por sua vez, sensibilizaria a alma. Essa, através da vontade, levaria a pequena glândula a mover-se, ativando as partes do sistema em direção às ações humanas. Esse tipo de solução, contudo, nunca foi inteiramente aceito: que tipo de interface seria a glândula pineal se ela teria de ser, *ao mesmo tempo*, algo físico e algo mental?

O problema colocado por Descartes parece não ter sido solucionado até hoje. O quebra-cabeças que dele herdamos é a necessidade de explicar como é possível a interação entre o físico e o mental. Voltaremos a falar desse problema e das soluções pro-

---

2. Refiro-me a Edmund Husserl.

3. Cf. Pinker (1997, p. 136), "Nós, humanos, podemos tomar uma proposição inteira e atribuir-lhe um papel em alguma proposição maior. Em seguida, podemos tomar a proposição maior e inseri-la em uma ainda maior... Assim como a capacidade de somar 1 a um número concede a capacidade de gerar uma série infinita de números, a capacidade de inserir uma proposição em outra concede a capacidade de ter um número infinito de pensamentos".

4. Cf., em especial, os artigos 32, 34 e 41.

postas pelos filósofos que sucederam Descartes mais adiante. Antes de empreender essa tarefa, analisaremos brevemente dois corolários que se seguem da separação entre mente e corpo formulada por Descartes: a inescrutabilidade dos estados subjetivos e a teoria cartesiana da natureza dos autômatos. Esses dois corolários ajudar-nos-ão a compreender melhor a verdadeira dimensão do dualismo cartesiano e as dificuldades a serem enfrentadas na formulação de teorias interacionistas. A inescrutabilidade dos estados subjetivos é retomada pela filosofia da mente contemporânea sob o nome de *teoria do acesso privilegiado* ou *acesso direto ao mental*, uma teoria que aprofunda ainda mais a existência de propriedades incompatíveis entre o físico e o mental. A teoria cartesiana dos autômatos reforça o dualismo substancial, sustentando a impossibilidade de replicação mecânica das atividades mentais humanas.

Acesso privilegiado e não espacialidade

Para Descartes a mente não poderia ser um objeto físico. Isto significa dizer que a mente não pode partilhar de nenhum tipo de propriedade que caracteriza esse último. Objetos físicos são extensos, isto é, ocupam lugar no espaço. Essa característica os torna sempre divisíveis, por oposição ao mental que é indivisível. Estados mentais não ocorrem no espaço, isto é, têm como característica a não espacialidade. A extensão ou espacialidade é uma propriedade inerente da matéria, mas não da mente. Da divisibilidade e da espacialidade, Descartes deriva mais uma característica dos objetos físicos: eles são *publicamente observáveis* por oposição aos estados mentais que são *privados*. Podemos, assim, representar a assimetria cartesiana entre mente e matéria ou entre *res cogitans* e *res extensa* na seguinte tabela[5]:

Objetos materiais	**Mentes**
Espaciais	Não espaciais
Extensos	Sem extensão
Públicos	Privadas

O que significa dizer que os estados mentais são privados? Significa que eles não são publicamente detectáveis. Um neurocirurgião que abrisse a cabeça de alguém e examinasse seu cérebro veria muitas células nervosas, mas nunca uma ideia. Mas isto nos revelaria apenas que os fenômenos mentais são invisíveis. Átomos também são invisíveis. Entretanto, seria difícil dizer que átomos são fenômenos mentais apenas porque não podemos observá-los. A diferença estaria no fato de que, apesar de os átomos serem invisíveis, nada impediria que um dia eles pudessem ser observados – por exemplo, através de um supermicroscópio eletrônico. O mesmo não ocorreria com os

5. Cf. HEIL, J. (1998, p. 18.)

fenômenos mentais: eles nunca poderiam ser observados, pois eles são privados. Dizer que eles são privados significa, antes de mais nada, dizer que eles são inescrutáveis: eles ocorrem *para mim*, ou seja, só eu posso saber o que estou pensando num determinado momento. Essa inescrutabilidade é a base para o chamado argumento do acesso privilegiado ou do acesso direto ao mental. Temos conhecimento imediato do que se passa em nossas mentes, mas o mesmo não ocorre com nossos próprios corpos: a mente é mais fácil de ser conhecida do que o corpo. Essa característica do mental seria mais um dos argumentos utilizados por Descartes para sustentar a existência de uma assimetria entre pensamento e matéria.

Segundo o argumento do acesso privilegiado, minha sensação de ter uma dor de dente me dá um conhecimento direto e imediato de que há algo errado com algum dente meu. Um dentista pode tirar um raio-x da minha boca e descobrir que tenho uma cárie nesse dente. Mas isto não permite que o dentista saiba que estou sentindo dor, a não ser que ele me pergunte, pois é possível ter um dente cariado e, mesmo assim, não sentir dor por um bom tempo. O raio-x não dá um acesso direto à sensação de dor, pois essa é intrinsecamente subjetiva e só eu poderia confirmá-la ou não.

O acesso privilegiado seria, ademais, infalível. Eu nunca poderia estar errado acerca daquilo que se passa na minha mente. Posso, contudo, estar errado acerca do que se passa no meu corpo. Comparemos dois enunciados: "Eu estou com dor de cabeça" e "Eu estou com febre". Só eu posso saber se estou com dor de cabeça ou não, minha sensação de dor apresenta-se como um conhecimento infalível que tenho acerca de meus próprios estados mentais. Não faria sentido se alguém me dissesse: "não, você não está com dor de cabeça agora". Posso fingir que estou sentindo dor, fazendo trejeitos, caretas e gritando, o que poderia levar outras pessoas a supor que estou com dores, mas, em última análise, só eu saberia se estou sentindo realmente alguma dor. É nesse sentido que a dor é um estado subjetivo inescrutável. O mesmo não ocorre com o segundo enunciado. Alguém poderia colocar um termômetro na minha boca e, após alguns minutos, afirmar: "não, você não está com febre". E essa pessoa estaria certa. A diferença estaria no fato de que a dor é uma sensação subjetiva, um estado mental, enquanto que "ter febre" é um estado do meu corpo – um estado ao qual não tenho acesso direto.

O acesso direto a meus próprios estados mentais e a infalibilidade desse tipo de conhecimento são o caminho para se sustentar que o conhecimento do corpo (e do cérebro) não implica num conhecimento da mente, sendo dois tipos diferentes de conhecimentos por se referirem a coisas (ou substâncias) distintas. Os filósofos da mente contemporâneos (pelo menos alguns deles) que sustentam o acesso privilegiado revivem um argumento que já era defendido por Descartes na sua tentativa de mostrar a existência de uma assimetria intransponível entre o físico e o mental.

Refutar os argumentos cartesianos do acesso direto e da não espacialidade dos estados mentais é uma tarefa à qual se propuseram vários filósofos da mente do século XX. Sustentar que a mente é distinta do corpo porque essa é mais fácil de conhecer parece pressupor, desde o início, o que se quer demonstrar, ou seja, que estados subjetivos são privados e inescrutáveis. Seguindo essa linha de raciocínio, o filósofo inglês

B. Williams (1978) sugere ainda um outro tipo de argumento para refutar as teses cartesianas. Ele nos diz que do fato de minha mente ser mais fácil de conhecer do que meu corpo não se pode inferir que essa deva, necessariamente, ser distinta do corpo, e que uma análise mais cuidadosa mostra que o argumento cartesiano apoia-se num falso raciocínio. Ele nos convida a imaginar que esse tipo de argumento equivaleria, numa disputa entre um dualista e um defensor da identidade entre mente e cérebro (identitarista), ao seguinte tipo de discussão, travado num baile de máscaras:

Dualista: Caro senhor "identitarista", bem à nossa frente encontra-se aquele homem mascarado cuja identidade eu não conheço. Mas eu conheço a identidade de meu pai; logo, o homem mascarado não é meu pai.

"*Identitarista*": Isso não é correto: é bem possível que seu pai seja o homem mascarado. Você só pode afirmar tal distinção depois que o homem mascarado tirar a máscara e revelar não ser o seu pai. E, cá comigo, afirmo que, depois de ser desmascarado, aquele homem revelará ser o seu pai.

Colocado nesses termos, o argumento de Descartes equivaleria a dizer que porque não conheço meu corpo (o homem mascarado que não conheço) tão bem como posso conhecer minha alma (a identidade de meu pai) daí se segue que minha alma é necessariamente distinta de meu corpo (logo, o homem mascarado não é meu pai). Esse é um argumento incorreto, que não resiste a um mínimo de análise e de reflexão. Por outro lado, o identitarista aposta que o "desmascaramento" do homem mascarado revelará que ele é o pai do Dualista ou que as evidências empíricas acabarão, um dia, por revelar que eventos mentais são eventos físicos[6].

Para refutar os argumentos cartesianos seria ainda preciso mostrar que estados mentais não são necessariamente inextensos e inescrutáveis. Discutiremos isto mais adiante. Antes de enveredar por essa tarefa mais complicada analisaremos, conforme prometemos, a teoria cartesiana acerca da natureza dos autômatos.

A teoria dos autômatos

A doutrina de Descartes inaugura o que ficou conhecido como *dualismo substancial* entre mente (*res cogitans*) e corpo (*res extensa*). O dualismo cartesiano – que talvez tenha aparecido na sua filosofia unicamente como uma etapa na sua busca por uma certeza primeira, inabalável, onde se poderia ancorar todo conhecimento humano – torna-se ainda mais problemático quando Descartes reflete sobre as semelhanças e diferenças entre autômatos, animais e seres humanos[7]. Parece que aqui encontramos

---

6. Esta crítica foi proposta por Williams (1978) e desenvolvida na dissertação de mestrado de Fábio C. Hansen (1995), sob minha orientação, na Universidade Federal de São Carlos.

7. A esse respeito poderíamos citar algumas passagens do *Discours de la Méthode* [Discurso do método] de Descartes (1637/1963). Mais ilustrativa, contudo, é a carta de Descartes ao Marquês de Newcastle, de 23 de novembro de 1646, onde suas posições são sustentadas de maneira mais explícita.

uma sutileza no pensamento cartesiano; uma sutileza que nos forçaria mais uma vez a aceitar sua proposta de uma distinção entre mente e corpo. Replicar o corpo seria condição necessária, mas não suficiente, para replicar um ser humano: autômatos, por mais perfeitos que fossem, nunca teriam estados mentais subjetivos e inescrutáveis. Nós, seres humanos, seríamos uma exceção ao materialismo e ao mecanicismo.

Já no século XVII, Descartes prenunciava o que seria o desenvolvimento futuro da robótica e da inteligência artificial e que tipo de consequências esse poderia ter para a sua filosofia. Esse era um tema que causava inquietação nos filósofos de sua época: haveria limites para o mecanicismo? Se a metáfora dos mecanismos e relógios era apropriada para descrever a natureza e os animais, seria ela adequada para explicar a natureza humana?

Um primeiro ensaio contendo reflexões sobre a natureza e potencialidades dos autômatos foi publicado em 1589 por Bernardino Baldi, abade de Guastalla. Nessa época, já havia alguns autômatos, todos eles bonecos de corda que imitavam, de uma maneira ou de outra, algumas atividades humanas. A construção de robôs – ou de autômatos, como eram chamados nessa época – viria a florescer no século XVIII. Alguns desses robôs tornaram-se célebres, como foi o caso, por exemplo, do pato de Vaucanson, que se supõe ter sido construído por volta de 1750, embora nunca se tenha sabido ao certo se eles de fato foram construídos ou apenas projetados.

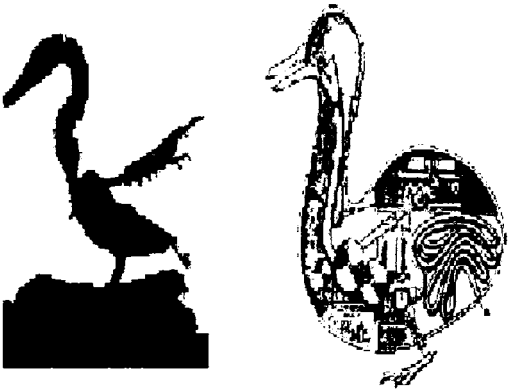

Fig. 2.1 – Duas possíveis reproduções do projeto do pato de Vaucanson. A primeira foi adaptada de Boden (1996). A segunda foi adaptada de Losano (1990), de onde retiramos grande parte do material histórico apresentado aqui.

Vaucanson sustentava que seu pato era um autômato capaz de reproduzir a atividade biológica desse tipo de animal. O pato seria capaz de esticar o pescoço, bicar um grão, engoli-lo, digeri-lo e expeli-lo – uma verdadeira imitação da função das vísceras destinadas à digestão. Vaucanson, contudo, nunca explicou como a digestão ocorria – ou ocorreria.

Outros autômatos famosos de que se tem notícia teriam sido os três bonecos construídos pelos irmãos Droz na Suíça, por volta de 1733. Um deles seria um escrivão, outro um desenhista e o terceiro uma tocadora. O escrivão seria capaz de escrever frases sobre uma folha de papel, o desenhista era capaz de fazer pelo menos cinco tipos de desenhos e a tocadora capaz de executar cinco melodias diferentes. Além disso, eles eram capazes de mover os cílios, os braços e inclinar o peito para a frente. Há registros também de uma máquina capaz de jogar xadrez, que teria sido construída pelo barão von Kempelen, no final do século XVIII, à qual, muitos anos mais tarde, Edgar Allan Poe teria dedicado um de seus contos.

A preocupação de Descartes era perfeitamente justificada: a exiguidade de recursos tecnológicos na sua época não implicava, necessariamente, que o projeto de replicar mecanicamente um ser humano fosse inexequível. E se acontecesse algum dia, mesmo que num futuro remoto, isto significaria que seres humanos seriam pouco mais do que sofisticados dispositivos de relojoaria.

Ora, essa possibilidade colidia frontalmente com o dualismo substancial. O pensamento não poderia ser o resultado de algum tipo de arranjo material entre peças ou de algum tipo de disposição orgânica, pois ele não é *res extensa*[8]. Os animais e os autômatos são desprovidos de pensamento e isto significa muito mais do que a impossibilidade de apresentar capacidades lógico-argumentativas, significa, sobretudo, a impossibilidade de possuir qualquer forma de estados ou processos conscientes. Os animais nada mais seriam do que organizações da substância extensa (*res extensa*) que, por mais complexas que fossem, poderiam ser explicadas em termos estritamente mecânicos. Assim como um pêndulo simples é mecanicamente menos complicado do que um relógio, alguns animais são mecanicamente menos complicados do que outros[9].

Para sustentar a existência de uma diferença qualitativa e intransponível entre humanos de um lado e animais e autômatos de outro, Descartes propõe dois testes: o teste da ação e o teste da linguagem. Embora nunca tenham sido explicitamente formulados, esses testes figuram em várias passagens de uma obra sua, publicada em 1637, o *Discurso do método*.

Descartes sustentava que entre humanos e animais há um paralelismo de estrutura fisiológica. Assim sendo, os animais teriam condições orgânicas para passar no teste da linguagem e, se não fazem isto, não se deve ao fato de lhes faltarem órgãos: os papagaios, por exemplo, podem proferir palavras como nós, embora não o façam testemu-

8. Nossa tendência habitual é concordar com Descartes e supor que o pensamento não tem a característica central da matéria, qual seja, a extensão ou espacialidade. Mas talvez esta seja uma falsa percepção que temos de nossos próprios processos mentais. Nos últimos textos de S. Freud, reunidos sob o título de *Escritos breves* – 1937-1938, há uma série de notas intituladas "Conclusões, ideias, problemas". Na nota de 22 de agosto de 1938, Freud nos diz que "A espacialidade é talvez a projeção do caráter extenso do aparelho psíquico. Nenhuma outra derivação é verossímil. Em vez das condições *a priori* de Kant, nosso aparelho psíquico [...] *psiché é extensa, mas nós não percebemos isto*" (ênfase nossa).

9. Reporto-me novamente à dissertação "Inteligência artificial e o problema mente-corpo" de Fábio C. Hansen (1995).

nhando que pensam o que dizem. Contudo, papagaios e homens têm um desempenho linguístico distinto – e isto não se deve à sua fisiologia. O papagaio, por mais que pudesse imitar nossa linguagem, nunca saberia *o que* está falando. Poderíamos treinar um papagaio para recitar um teorema matemático com perfeição. Ele não saberia tampouco *do que* estaria falando, ou seja, ele não estaria expressando nada, nenhum tipo de significado. A linguagem seria mais do que simplesmente a emissão de sons em código e a habilidade física necessária para fazê-la. Poderíamos construir autômatos ou mecanismos capazes de proferir palavras e sentenças, mas eles estariam na mesma situação do papagaio.

O teste da ação segue o mesmo raciocínio: um autômato poderia executar qualquer tipo de ação feita por um ser humano. Poderíamos construir um autômato com uma enorme quantidade de programas e de sub-rotinas, cada uma delas correspondendo a um tipo de tarefa normalmente executada por seres humanos. Algumas ações poderiam ser executadas com mais exatidão e mais eficiência pelo autômato, mas seriam apenas comportamentos automáticos, desprovidos de consciência. Mas o autômato nunca saberia o que está fazendo, ele seria, no máximo, um zumbi bem construído. Ele não estaria *agindo* verdadeiramente, mas apenas executando programas ou sub-rotinas. Como consequência, o autômato não saberia quando e por que passar de um programa ou uma sub-rotina para outra quando fosse necessário, ou seja, ele não teria a flexibilidade necessária para *variar* o curso de suas ações, isto é, quando lançar mão de um programa em vez de outro. A ele faltaria uma *racionalidade universal* que seria típica dos seres humanos e não seria replicável em termos materiais[10]. A racionalidade universal implicaria em consciência e essa tampouco seria reprodutível por algum tipo de arranjo material. Isto faria com que a autonomia do autômato – se é que podemos fazer esse trocadilho – fosse sempre muito restrita.

Ora, seria possível superar esse entrave colocado pelo teste da ação? Uma resposta preliminar poderia ser positiva: poderíamos construir um autômato que incorporasse um número imenso de sub-rotinas. Mas isto não seria suficiente para invalidar o teste da ação. *Quando* e *por que* passar de uma sub-rotina para outra continuaria sendo um problema. A ausência de uma racionalidade universal ainda nos forçaria a postular a existência de um hiato intransponível entre autômatos e seres humanos. O *corpo* humano poderia ser perfeitamente replicado na construção de um autômato. Uma réplica extremamente sofisticada e aperfeiçoada poderia – no limite – exibir comportamentos indistinguíveis daqueles de quem a construiu, mas, mesmo assim, a ela faltaria a consciência. Em outras palavras, replicar as características físicas do ser humano seria condição *necessária*, mas não *suficiente* para replicar nossa vida mental. A um autômato que pudesse falar como nós, executar operações matemáticas abstratas ou exibir comportamentos complexos faltaria ainda a *res cogitans*, algo como uma quintessência subjacente a todos os seus comportamentos manifestos que os tornaria conscientes.

10. A mesma observação é feita por REY, G. (1997, p. 43).

Quando a glândula pineal dessas aves é extraída de seus cérebros, nota-se que, ainda assim, ela é sensível à luz. O mais interessante é que a glândula pineal reage especificamente quando se passa do escuro para a luminosidade: com certeza é isto que faz com que o galo acorde e depois cante. Isto ocorre porque a glândula pineal produz um hormônio específico, a melatonina. Níveis maiores ou menores de melatonina encontram-se no cérebro, dependendo dos períodos do dia. Sabe-se também que, se injetarmos melatonina num pássaro, em poucos minutos ele cairá em sono profundo[11].

Assim sendo, a hipótese de Descartes não era completamente insensata. A glândula pineal relaciona-se com consciência, se entendemos essa palavra como sinônimo de vigília e por oposição ao estado de inconsciência no qual estaríamos submersos durante nossos períodos de sono. Certamente Descartes não poderia saber nada acerca das funções cerebrais da glândula pineal na sua época, embora sua hipótese revele-nos uma intuição incomum. Uma intuição que, contudo, não foi capaz de resolver satisfatoriamente como se daria a ligação entre mente e corpo, pois, para a glândula pineal poder oferecer a interface desejada, teria de ser algo intermediário entre o físico e o mental, o que se torna inconcebível à medida que sabemos que ela é parte do cérebro. Essa foi, pelo menos, a maneira como seus sucessores interpretaram a teoria da glândula pineal e a consideraram insatisfatória.

O resultado dessa insatisfação foi a parafernália teórica que constitui a história das soluções possíveis para o problema cartesiano. Teorias bizarras apareceram como tentativas de solucionar esse problema, como por exemplo a proposta da existência de uma harmonia preestabelecida entre mente e corpo (formulada por Leibniz) ou o ocasionalismo (formulado por Malebranche). Leibniz sustentou a possibilidade de existir um paralelismo psicofísico embora não existisse nenhum contato causal entre o físico e o mental. A sucessão de eventos físicos e a sucessão de eventos mentais ocorreria em paralelo no tempo, numa extraordinária coincidência de simultaneidades que seria garantida por Deus. Malebranche sustentou que esses dois processos, o físico e o mental, seriam independentes, mas correlacionados pela interferência divina. Outros filósofos, como Hobbes e La Mettrie, preferiram sustentar um monismo materialista por oposição ao dualismo cartesiano.

A herança cartesiana de um problema não resolvido – talvez insolúvel nos termos em que foi formulado – e de uma série de soluções consideradas insatisfatórias atravessou não apenas a história da filosofia como também propagou-se, de modo implícito, para outras disciplinas que apareceram no final do século XVIII, como é o caso, por exemplo, da psiquiatria. Se percorrermos os principais tratados legados pelo seu fundador, verificaremos que, curiosamente, ao tentar definir a "doença mental", Pinel oscilou em situá-la seja no âmbito da mente, seja no âmbito do corpo[12]. Se os distúrbios identificados por Pinel devessem ser considerados uma "doença" seria melhor si-

11. A esse respeito, cf. Greenfield (1997, p. 60.)

12. Cf. Pinel (1801).

tuá-los no âmbito da medicina e, portanto, no âmbito do corpo. Por outro lado, a ideia de doença *mental* exigia que nela houvesse algum componente que não poderia ser reduzido inteiramente ao domínio do físico.

Vários anos mais tarde, quando a psicologia se separou da filosofia para tentar tornar-se uma disciplina independente, verificamos que a oscilação entre o físico e o mental foi também inevitavelmente herdada pela nova ciência da mente que tentava nascer. Seu fundador, W. Wundt (1832-1920), embora propondo que a psicologia deveria ser o estudo científico da consciência, e sustentar a independência dessa em relação a sua base neural, não pôde ignorar os desafios e problemas resultantes da adoção de um paralelismo psicofísico. O mesmo tipo de dificuldade parece atravessar quase todas as teorias psicológicas contemporâneas.

Curiosamente, o desenvolvimento da psiquiatria e, mais tarde, da psicologia, parece ter ignorado os trabalhos do filósofo alemão I. Kant (1724-1804) que, no seu livro, *Kritik der Reinen Vernunft* [Crítica da razão pura], publicado em 1781, sugeria que o problema das relações entre mente e corpo seria insolúvel. A razão disto talvez se deva ao fato de sua obra apresentar uma extraordinária dificuldade de compreensão e de interpretação. Kant operou uma revolução na filosofia ao afirmar que embora existam objetos reais no mundo só podemos enxergá-los e concebê-los com a forma com que nossa mente os vê ou os concebe. Em outras palavras, nossa experiência é inevitavelmente modelada pelo nosso aparelho cognitivo e nunca poderemos saber se o mundo é aquilo que se apresenta a nós ou se é algo diferente[13]. O mundo que nosso aparelho cognitivo nos apresenta constitui aquilo que ele chamou de "limites da experiência possível". Contrariamente a Descartes, Kant não acreditava em poderes ilimitados da razão humana e sustentou que existem limites para aquilo que podemos conhecer. Esses limites são dados pelo âmbito da "experiência possível". Quando se esquece isso, passa-se a fazer um uso ilegítimo da razão e é desse uso ilegítimo que deriva a maioria dos erros da filosofia.

As tentativas de resolver o problema mente-corpo seriam, no entender de Kant, um caso típico de uso ilegítimo da razão – um erro que resulta de ignorar que nosso conhecimento está confinado aos limites da experiência possível. Quando sustentamos tanto que mente e cérebro são a mesma coisa, quanto que não o são, esquecemos que não podemos falar nem de mentes e nem de cérebros, mas apenas de como mentes e cérebros *se apresentam para nós*, pois todas as nossas experiências são modeladas por nosso aparelho cognitivo, inclusive nossas percepções de mentes e de cérebros. Nunca teríamos acesso a um cérebro, mas unicamente a algum tipo de representação do cérebro produzida pelo nosso aparelho cognitivo. Assim, quando afirmamos a equação *estados cerebrais = estados mentais*, estamos apenas formulando um pensamento. Um pensamento acerca do cérebro que o relacionaria a outro pensamento, acerca da mente. Nunca teríamos condição de saber se essa equação é verdadeira ou não, pois isto

13. Obviamente esta é uma supersimplificação pela qual pedimos desculpas aos filósofos acadêmicos.

exigiria que transcendêssemos nossos próprios pensamentos. Só poderíamos verificar a verdade ou não dessa equação se pudéssemos nos situar fora de nossos próprios pensamentos, o que seria impossível: não podemos ultrapassar os limites da nossa experiência possível. Tentar resolver esse problema seria supor que nossa razão e nosso conhecimento poderiam ultrapassar-se a si próprios; uma ambição natural de nossa razão que nunca poderia ser realizada. Alguns filósofos da mente contemporâneos, como Colin McGinn, sustentam, hoje em dia, um ponto de vista parecido, embora curiosamente pareçam esquecer-se de fazer qualquer referência a esse argumento que Kant formulou há mais de 200 anos.

Na sua interminável oscilação pendular, o desenvolvimento da psicologia e da filosofia da mente no decorrer do século XIX enveredou por tentativas materialistas de vários tipos. O dualismo, que aparentemente teria triunfado na história da filosofia posterior a Descartes, começou, gradualmente, a ceder espaço para teorias monistas. Há várias razões para explicar esse tipo de mudança, mas talvez uma das mais importantes tenha sido a teoria da evolução formulada por Darwin, da qual se tentou derivar algum tipo de teoria biológica e evolucionária da mente. Essa tarefa teria sido empreendida pelo próprio Darwin que sustentou a existência de uma continuidade entre os processos mentais dos seres humanos e os dos animais. A mente, assim como outras características físicas dos seres humanos, seria resultado de um processo evolucionário guiado pela seleção natural e pela adaptação ao meio ambiente. Num de seus trabalhos iniciais, só recentemente reeditado na coletânea de P. Barrett e H. Gruber, *Metaphysics, Materialism and the Evolution of Mind* [Metafísica, materialismo e a evolução da mente] (1974), Darwin lançou os princípios daquilo que hoje constitui a chamada psicologia evolucionária, ou seja, a ideia de que nossa vida mental, nossas emoções e até mesmo nossa organização social derivam-se de nossos instintos básicos e das vicissitudes que esses teriam sofrido no decorrer da evolução e da seleção natural.

Paralelamente a essa tendência de caráter monista, surgem no final do século XIX fortes reações às teorias biológicas da mente ou teorias naturalistas. A história da filosofia inicia uma nova oscilação pendular, dessa vez em direção ao dualismo. Filósofos como Franz Brentano (1838-1917) e Edmund Husserl (1859-1938) escreveram contra a plausibilidade das teorias naturalistas da mente e do monismo materialista.

Brentano, na sua obra *Psychologie von empirischen Standpunkt* [Psicologia de um ponto de vista empírico] (1925), sustentou a existência de uma marca distintiva do mental, irredutível a qualquer substrato físico. Essa marca distintiva ele chamou de *intencionalidade*. A palavra intencionalidade deriva-se de "intenção" (*intentio*) um termo que, normalmente, associamos com nossas ações: as intenções imprimem uma direção para nossas ações. O mesmo ocorre também com meus desejos. Não existem intenções e desejos sem conteúdo: esses são sempre *acerca de alguma coisa* e é isto que caracteriza sua direcionalidade ou intencionalidade.

Brentano sustentou que todos os estados mentais têm essa marca distintiva. Consideremos, por exemplo, nossas percepções visuais. Quando olho para uma lâmpada ace-

sa, certamente ondas luminosas ou fótons estão chegando à minha retina. Minha retina transmite esse tipo de informação para o meu cérebro que a processa – um cérebro que, por sua vez, está no interior da caixa óssea que compõe meu crânio. Contudo, vejo a lâmpada como se ela estivesse fora de mim. Meus estados mentais *referem-se* ou *direcionam-se* a algo que está fora de mim, ou seja, meus estados mentais *são sempre acerca de algo fora de mim*. Essa referência implícita a algo fora de mim é a marca intencional de meus estados mentais. É através dessa marca intencional que meus estados mentais adquirem *significado*, ou seja, passam a ser *representações* de coisas no mundo.

Os objetos a que se referem meus estados mentais não precisam existir concretamente: eles são objetos *possíveis*. Minha linguagem – um complexo e sofisticado sistema de representações também dotado de intencionalidade – é um bom exemplo disso. Posso construir sentenças acerca de fatos ou objetos que não existem ou nunca existiram, mas essas sentenças poderão ter, contudo, significado. O mesmo pode ocorrer com meus pensamentos: posso pensar num unicórnio e esse pensamento pode ter significado para mim, ou seja, ele será um estado mental dotado de intencionalidade, embora não corresponda concretamente a nenhum objeto no mundo.

Brentano sustentou que essa marca intencional – ou essa direcionalidade do pensamento – estabelece uma separação entre o físico e o mental por ser uma relação de um tipo especial, distinta daquela que ocorre entre objetos físicos. Representar um objeto possível, ou o *direcionar-se para algo* característico de nossos pensamentos que constitui a relação intencional exigiria sempre uma mente imaterial. Décadas mais tarde, o filósofo norte-americano John Searle retoma o tema da intencionalidade para sustentar que nunca poderíamos replicar nossa vida mental usando computadores, porque a esses sempre faltaria esta marca distintiva do mental, qual seja, a intencionalidade. Searle não defendeu a imaterialidade da mente a partir da intencionalidade, mas a ideia de que a relação entre símbolos operada pelos computadores digitais seria insuficiente para gerar uma relação intencional autêntica. Examinaremos alguns aspectos dessa polêmica no capítulo VI quando discutirmos o chamado argumento do quarto do chinês.

Searle não foi o único filósofo da mente contemporâneo a retomar o tema da intencionalidade do mental. Tentativas de conceber ou de reduzir a relação intencional a algo físico ou biológico povoam o cenário da filosofia da mente no século XX. Esta tem buscado na física e na neurociência elementos e argumentos para refutar o dualismo, reiniciando uma aposta na perspectiva materialista ou, pelo menos, uma nova oscilação pendular na história da filosofia.

*O século XX: teoria da relatividade

O dualismo de substâncias proposto por Descartes – o dualismo que divide o mundo em *res cogitans* e *res extensa* – é, hoje em dia, considerado por muitos como sendo uma doutrina obsoleta. O século XX é considerado por muitos como trazendo a marca do triunfo do monismo materialista como opção teórica de quase todos os filósofos

deste século. O desaparecimento gradual do dualismo, entretanto, não tem ocorrido sem deixar sequelas dos mais variados tipos, como, por exemplo, a contradição constante entre o materialismo como opção científica e filosófica com as crenças religiosas que professamos. Algumas características de nosso vocabulário cotidiano ainda se referem, mesmo que implicitamente, a uma divisão do mundo entre substância física e substância mental. Elas são sinais de que não nos livraremos tão facilmente da herança cartesiana.

Que tipo de razões, descobertas ou argumentos a ciência nos proporcionou no ao longe do século XX para abandonarmos o dualismo de substâncias? Em primeiro lugar, tivemos o desenvolvimento da física contemporânea, em especial a teoria da relatividade. Em segundo lugar, os avanços da neurociência e a aposta na possibilidade de explicar a natureza dos fenômenos mentais, incluindo a própria consciência, a partir da investigação do funcionamento de nosso cérebro. A identidade entre mente e cérebro parece ter se configurado progressivamente como uma espécie de pressuposição tácita dos neurocientistas, uma espécie de *ideologia científica* ou *filosofia espontânea*, que produz irritação quando questionada. Contudo, a pressuposição de uma identidade entre mente e cérebro é ainda uma hipótese que tem resistido a qualquer tipo de confirmação empírica conclusiva, constituindo uma enorme extrapolação baseada em alguns dados experimentais.

Examinaremos a seguir em que sentido a teoria da relatividade força, no limite, o abandono definitivo do dualismo de substâncias. Na próxima seção apresentaremos um experimento realizado por neurocientistas na década de 1970 – um experimento que, aparentemente, poderia nos levar a desacreditar uma das principais teses cartesianas, qual seja, a de que a mente, ao contrário da matéria, seria essencialmente indivisível. Verificaremos que esse experimento está ainda longe de oferecer evidências cruciais ou decisivas para sabermos se Descartes estava certo ou não. Como todos os experimentos em neurociência e em psicologia, os problemas não surgem da existência ou não de dados empíricos, mas, fundamentalmente, das controvérsias inevitavelmente envolvidas na sua interpretação.

Em que sentido a teoria da relatividade abala o dualismo de substâncias? Retomemos por um momento os argumentos de Descartes em defesa do dualismo. Vimos que um de seus argumentos, o da indivisibilidade do pensamento (*res cogitans*) por oposição à matéria (*res extensa*), implica a não espacialidade do pensamento e, desta última, segue-se a não localidade do mental. Aparentemente, Descartes identificava estados mentais com estados conscientes. Estados conscientes não são percebidos da mesma maneira que percebemos coisas à nossa volta[14]. Percebemos vários objetos no espaço através da visão e é a visão desses objetos que faz com que os percebamos como coisas no espaço. Contudo, não percebemos nossos estados mentais como coisas no espaço, pois nós

14. Cf. McGinn (1995).

não podemos tocá-los, cheirá-los e assim por diante. Não observamos nossos estados mentais da mesma maneira que observamos algo físico no espaço, como, por exemplo, um pássaro voando ou uma descarga elétrica. Essa peculiaridade do mental em contraste com o físico não surge apenas na filosofia cartesiana. Mais recentemente, o filósofo austríaco Ludwig Wittgenstein sustentou o mesmo ponto de vista[15].

Ora, se estados mentais não ocorrem no espaço, eles ocorrem, porém, no tempo. A sucessão de meus estados mentais forma uma sequência discernível. O fluxo temporal parece ser uma característica constitutiva da nossa consciência, que percebemos como uma constante metamorfose do presente em direção ao futuro. Esse tipo de percepção do tempo, a partir da qual organizamos nossa experiência cotidiana, corresponde à concepção clássica de ordem temporal que encontramos na mecânica newtoniana. O tempo teria uma marcha única, objetiva e universal.

Essa imagem do tempo é radicalmente alterada pelo advento da teoria da relatividade. O tempo deixa de ser considerado um parâmetro abstrato e universal e passa a ser uma dimensão da realidade física, a quarta dimensão dos objetos. Presente, passado e futuro deixam de ser definidos a partir de uma marcha universal do tempo e passam a depender da posição e do movimento do observador – de um observador situado em algum lugar do espaço. O que entendemos por "presente" passa a depender de uma escolha arbitrária de um sistema de coordenadas[16].

É a adoção de um sistema de coordenadas – minha posição no espaço, por exemplo – que determina o que eu perceberei como sendo um evento no passado, no presente ou no futuro. De uma maneira muito simplificada poderíamos dizer o seguinte: a luz do sol que percebemos agora, neste momento, é um evento que ocorreu oito minutos atrás. Oito minutos é o tempo que essa luz leva para viajar do sol até o planeta Terra e incidir sobre minha retina. Quando olhamos para o céu estrelado, de noite, o que vemos é, em sua maior parte, uma fatia do *passado* do universo. Pois a luz colorida das estrelas que chega aos nossos olhos teve de percorrer uma longa distância até chegar a nós. O que vemos já ocorreu; o que vemos já é parte do passado do universo. Se uma dessas estrelas explodisse, continuaríamos, ainda, por um bom tempo, a vê-la como a víamos antes, pois até que a luz da explosão chegasse a nós levaria um bom tempo. Ora, o que significa esse "levaria um bom tempo"? Significa a *distância* que essa luz tem de percorrer no espaço. Se estivéssemos em outra posição no espaço, muito possivelmente esses intervalos de tempo seriam diferentes. Mais do que isto: nossa posição no espaço é que determina, em última análise, nossa própria *ordenação* dos eventos em passado, presente e futuro. Por causa de minha posição no espaço vejo primeiro a luz do sol e só muito tempo depois eu veria a explosão da estrela. Talvez eu nem sequer sobreviva para chegar a ver essa explosão. De qualquer forma, se eu vivesse, *primeiro* veria a luz do sol e *depois* a explosão da estrela. Para um outro observador, situado

15. Cf. Wittgenstein (1969, p. 7 e 8).

16. Para uma descrição detalhada, cf. Lockwood (1989) e Russell (1925).

numa outra região do espaço, poderia ocorrer o contrário: ele *primeiro* veria a explosão da estrela e *depois* a luz do sol. O "antes" e o "depois" tornam-se relativos à situação do observador no espaço. É isto que significa dizer que o tempo é uma quarta dimensão dos objetos físicos: tempo e ordenação temporal, sequencialidade, dependem da posição do observador no espaço e, nesse sentido, não podemos falar de tempo isoladamente, mas apenas de espaço-tempo.

Ora, essa concepção de espaço-tempo proposta pela teoria da relatividade implica que na verdade não existe passado, presente e futuro, mas apenas que fazemos um recorte na sequência de eventos físicos do universo, um recorte a partir de nossa posição como observador e que, a partir dela, *construímos* uma sequencialidade que ordena esses eventos de acordo com nossa percepção. Há uma diferença entre o modo como eventos ocorrem e se sucedem no universo e o modo como eles ocorrem e se sucedem *para nós*. A ordenação de tempo que fazemos, indo de um passado em direção a um presente e um futuro, numa sequência linearizada, é uma construção feita pela nossa consciência. A ela não corresponde nenhum tipo de realidade física a não ser nossa posição no espaço a partir da qual fazemos esse recorte entre passado, presente e futuro.

Ora, se não existe tempo independentemente do espaço e se estados mentais ocorrem no tempo, eles têm de, necessariamente, ocorrer no espaço também. É sobre esse ponto que Lockwood (1989) nos chama a atenção para enfatizar, então, que a aceitação da teoria da relatividade força-nos, igualmente, a aceitar algum tipo de identidade entre estados mentais e estados físicos. Pois se estados mentais ocorrem no tempo implica que eles têm de ocorrer no espaço. Isto implica, por sua vez, que eles teriam igualmente de ser algum tipo de estado físico, pois a espacialidade é característica daquilo que é físico.

Claro que alguém poderia objetar que aqui estaríamos incorrendo numa circularidade, ou seja, que tomamos como ponto de partida que estados mentais são algo físico, para então dizer que eles ocorrem no tempo e, portanto, de acordo com a teoria da relatividade, eles têm de ser eventos espaciais também. Essa objeção, contudo, é incorreta. Não precisamos pressupor que estados mentais sejam, logo de início, algo físico, para afirmar que eles ocorrem no tempo e, depois, derivar a consequência de que eles tem de ocorrer no espaço. O fato de eventos mentais ocorrerem no tempo é um dado imediato: seria inconcebível não apreendê-los como formando uma sequência – ou pelo menos *algum tipo* de sequência, mesmo que essa seja dependente apenas da posição do observador no espaço. Ou seja, pouco importaria se a sequência que damos a eles é uma construção da nossa consciência, pois mesmo numa construção estaria pressuposto o espaço-tempo e não um tempo universal. De qualquer maneira, estaria envolvida uma dimensão física nessa sucessão, uma dimensão física que nos forçaria a admitir, pelo menos, que eventos mentais são alguma *coisa* no mundo, ou seja, parte da realidade física. Restaria então saber se essa realidade física necessariamente corresponde a algum tipo de realidade cerebral, ou seja, se estados mentais ocorrem especificamente *no cérebro*, para podermos, então, identificá-los com eventos cerebrais.

Isso significa que, para passar da ideia de que estados mentais são estados físicos e que estados físicos são estados cerebrais eu preciso pelo menos garantir que esses estados mentais ocorrem *no* cérebro. Mas garantir que estados mentais ocorrem no cérebro e, por ocorrerem em regiões específicas desse, perfeitamente delimitáveis, para, em seguida, propor uma identidade entre estados mentais e estados cerebrais é um passo maior que talvez não possa ser derivado da proposta de Lockwood. De sua proposta podemos apenas derivar que meus pensamentos ocorrem no meu corpo, mas não que eles ocorrem especificamente no meu cérebro. Pois é a posição de meu corpo no espaço que determina a sequencialidade dos meus pensamentos e sua espacialidade como resultante do espaço-tempo. Meus pensamentos estariam no meu cérebro unicamente porque meu cérebro é parte do meu corpo. Mas essa é uma localização vaga, bastante imprecisa, que nos permite unicamente dizer que *possivelmente* meus pensamentos ocorrem na minha cabeça e não nas minhas pernas.

Em primeiro lugar preciso saber onde está meu corpo. Saber onde está meu corpo é algo que sei a partir de um conjunto de relações causais que presumo existir entre objetos que estão à minha volta e a ocorrência dessas percepções na minha cabeça. Ou seja, localizo meu corpo a partir de relações sensoriais que ele mantém com os objetos que estão à sua volta. Se vejo a Torre Eiffel é porque meus olhos – e portanto meu corpo – estão perto da Torre Eiffel. É bem provável então que meu pensamento ou minha percepção da Torre Eiffel enquanto estado mental esteja também perto da Torre Eiffel. Mas esse é um modo aproximado e indireto de dizer onde está meu estado mental correspondente à percepção da Torre Eiffel. Esse modo aproximado e indireto me permite afirmar que a percepção da Torre Eiffel ocorre no meu cérebro. Não em algum lugar *específico* do meu cérebro, mas, muito possivelmente, em algum lugar da minha cabeça unicamente porque essa normalmente está onde meus olhos estão.

Mas como posso saber onde ocorre meu pensamento de que $2 + 2 = 4$? Nesse caso não tenho nenhum tipo de ligação sensorial (através de meus olhos) com algum objeto no mundo como é o caso do estado mental correspondente à percepção da Torre Eiffel. Não posso ter nenhuma ligação sensorial com o meu próprio cérebro, da mesma forma que, por exemplo, meus olhos podem ter uma ligação sensorial com a Torre Eiffel. O tecido cerebral é insensível, isto é, o cérebro não tem sensações de si mesmo ou sensações do que estaria ocorrendo nele, o que me impede de saber que estou pensando com meu cérebro, isto é, que os pensamentos ocorrem na minha cabeça. Em outras palavras, não posso estabelecer com meu próprio cérebro uma relação sensorial parecida com aquela que estabeleço com a Torre Eiffel quando estou perto dela – uma relação sensorial que me permite supor, adequadamente, que estou de fato perto da Torre Eiffel. Não posso *sentir* que meus pensamentos estão ocorrendo no meu cérebro; não há a possibilidade de se estabelecer uma ligação sensorial e causal que me permitiria saber que meus pensamentos estão ocorrendo no meu cérebro.

A proposta de Lockwood permite-nos concluir apenas que estados mentais são estados físicos, mas não necessariamente estados cerebrais. Pois se a teoria da relatividade força-nos a aceitar a ideia de que estados mentais, por ocorrerem no tempo, têm,

necessariamente, de ter a espacialidade que caracteriza as coisas físicas, isto é insuficiente para afirmar que eles ocorrem no cérebro ou em algum lugar específico deste. Não dispomos tampouco de evidências sensoriais de que nossos pensamentos ocorram no nosso cérebro, fazemos apenas algumas inferências aproximadas no caso de percepções de objetos à nossa volta. Entretanto, meu sentido introspectivo me revela como autoevidente que meu pensamento de que 2 + 2 = 4 ocorre no meu cérebro. Estaríamos então de volta ao primado da evidência introspectiva como queria Descartes? E de volta à ideia cartesiana de que a mente é mais fácil de conhecer do que o corpo, na medida em que essa nos fornece evidências introspectivas infalíveis? Para que o dualismo cartesiano pudesse ser inteiramente refutado – como quer Lockwood – seria preciso não só mostrar que o mental não pode ser imaterial, por ser necessariamente espacial, como também solapar a tese do acesso privilegiado (que tornaria a mente diferente da matéria na medida em que a primeira seria mais fácil de ser conhecida) ou da evidência introspectiva imediata. Em outras palavras, seria preciso mostrar que a evidência introspectiva estaria longe de ser infalível e de nos proporcionar um conhecimento imediato acerca de nossa própria mente.

A refutação completa do dualismo passa pela refutação da ideia de que nossas "evidências introspectivas" sejam certeiras e não possam nos conduzir a nenhum tipo de erro ou paradoxo. Nesse caso específico, abandonar a evidência introspectiva significa rejeitar a ideia de que o que chamamos de "pensamento" ocorra apenas no cérebro. O sistema nervoso, que parte do cérebro, espalha-se pelo corpo todo; a divisão que tendemos a fazer entre "cérebro" e "o resto do corpo" é convencional, da mesma forma que a divisão que se faz, para fins de estudo, entre esôfago e estômago, como se esses dois não estivessem conectados e um não fosse a continuação do outro. Se expandirmos dessa maneira nossa concepção de localização espacial do pensamento, desfaremos a evidência introspectiva sobre a qual se apoia a tese cartesiana do acesso privilegiado.

A questão da localização da dor, que tanto tem ocupado os neurocientistas, constitui um exemplo da necessidade desse tipo de expansão. Quando queimo minha mão, a dor ocorre na mão ou no cérebro? Algumas pessoas diriam que a dor ocorre na mão, mas, como isto seria possível sem um cérebro? Quando tomo anestesia interrompo as ligações entre a mão e o cérebro; alguém poderia cortar minha mão e eu não sentiria dor. Logo, a fonte da dor seria o cérebro e a dor *estaria* no cérebro. Mas, como o cérebro poderia produzir dor se seu tecido é insensível? Rapidamente chegamos ao mesmo tipo de paradoxo, ao tentar situar a dor ou na mão ou no cérebro, em vez de conceber que mão e cérebro estão interligados no sistema nervoso. Essa ideia de interligação explicaria, ademais, a natureza das chamadas "dores dos membros fantasmas" ou dores que algumas pessoas continuam a sentir meses após terem um de seus membros amputados.

A tese do acesso privilegiado não produziu apenas alguns paradoxos que, aparentemente, ameaçam a proposta de Lockwood, ao forçar-nos a aceitar a ideia de que existe um conhecimento introspectivo certeiro e infalível que se choca com o que teo-

rias científicas e neurológicas sustentam. Ela nos remete a um outro tipo de paradoxo que a tradição cartesiana gostaria de fazer passar desapercebido. Armou-se um outro tipo de armadilha, à qual parecem ter sucumbido muitos filósofos da mente posteriores a Descartes. Por ter tomado a evidência introspectiva como ponto de partida, deu-se mais um passo em falso: um passo em direção à ideia de que aquilo que ocorre no meu cérebro é algo interior a mim, como se meu pensamento pudesse *não estar ocorrendo* no mundo. Celebrou-se aquilo que alguns chamaram de "mito da interioridade". Esse é, sem dúvida, um mito, pois meu cérebro é parte do mundo, da mesma maneira que aquilo que está ocorrendo dentro dele, ou seja, meus pensamentos.

Alguns filósofos tentaram desmontar esse mito (veremos isto no capítulo V), sem, entretanto, explicar por que todos nós *experienciamos* esse mito. Da divisão arbitrária entre interior e exterior ou entre interno e externo que experienciamos introspectivamente surgiram outros paradoxos, quase que em efeito-dominó. Surgiu a possibilidade de separar mente, cérebro e corpo, pois a mente passa a poder referir-se a si própria como não sendo parte de um corpo ou de um cérebro. Ou como apenas *estando* num corpo ou num cérebro que poderia não lhe pertencer. Paralelamente surgiu o problema da intencionalidade que nada mais é do que outra versão da aposta na evidência introspectiva, levando, contudo, ao problema paradoxal de saber como é possível que aquilo que ocorre dentro de mim capte, ao mesmo tempo, algo que está sempre e inexoravelmente fora de mim.

Os paradoxos da localização de estados mentais – frequentemente apontados como uma das principais objeções ao materialismo – são, na verdade, uma consequência do mito da interioridade. Pois, para existir, o teatro introspectivo não pode estar em lugar algum, da mesma maneira que os elementos que o compõem, ou seja, nossos pensamentos. Montou-se uma grande parafernália teórica acerca da falta de sentido ou do aspecto paradoxal das sentenças que exprimiriam a localização de pensamentos no cérebro (voltaremos a esse assunto no próximo capítulo).

Mas a grande herança – a herança verdadeiramente problemática – deixada pelo cartesianismo na sua aposta no paradoxo da interioridade foi a divisão do mundo do conhecimento em duas partes: de um lado, a imagem do mundo e da mente vista pelo lado *subjetivo*, *interior*, e, de outro lado, a imagem do mundo e da mente vista pela ciência. Essas seriam imagens irreconciliáveis. Quando se estuda a mente ou se busca entender a natureza dos fenômenos mentais, essa divisão assume a forma de uma intransponibilidade entre experiência subjetiva e os dados e hipóteses fornecidos pela neurociência. A história de como se tem tentado, através dos mais variados artifícios teóricos, reconciliar essas duas imagens é a própria história da filosofia da mente nas últimas décadas.

*O século XX: a neurociência[17]

A possibilidade de desafiar a herança cartesiana no século XX foi, em grande parte, proporcionada pelo desenvolvimento da neurociência. Os neurocientistas, nas últimas décadas, têm persistentemente procurado soluções empíricas para os problemas da filosofia da mente – soluções que surgiriam de um estudo aprofundado do funcionamento do cérebro. Quase todos eles buscam o "desmascaramento" do qual falamos algumas páginas atrás; um "desmascaramento" que mostraria, finalmente, que os chamados eventos mentais são eventos físicos ou cerebrais.

Contudo, neurocientistas e filósofos da mente se desentenderam quanto à interpretação de alguns experimentos cruciais que, de acordo com os primeiros, revelariam que o dualismo substancial de Descartes estaria errado. Um exemplo típico desse tipo de experimento foram as *comissurotomias* (também chamadas de *calosotomias*) ou experiências com os chamados *split-brain patients*, ou seja, dividir cérebros para saber se, ao fazê-lo, dividimos também mentes. Se isto ocorresse, então o argumento cartesiano da assimetria entre o físico e o mental baseado na indivisibilidade deste último (proposto na sexta meditação) seria refutado.

Os experimentos a que nos referimos começaram na década de 1950 e tiveram como pioneiros Ronald E. Myers e Roger W. Sperry, que fizeram uma descoberta interessante: mostraram que, quando era seccionado o corpo caloso (uma das estruturas que une os hemisférios cerebrais) de gatos, cada metade hemisférica funcionava independentemente, como se fosse um cérebro completo. Executando a desconexão inter-hemisférica e a ruptura do quiasma óptico (ponto de cruzamento dos nervos ópticos), criava-se a condição para poder verificar como cada hemisfério lidava com as informações visuais que só a ele chegavam. As imagens apresentadas ao olho esquerdo eram conduzidas somente ao córtex visual esquerdo e, de modo idêntico, ao direito. Colocava-se o animal para trabalhar um problema com apenas um olho; o outro permanecia tampado. Quando esse primeiro olho era tampado e o problema apresentado ao segundo olho, o animal não apresentava reconhecimento do problema e tinha que aprender desde o início com o outro hemisfério.

Esse achado levou a um intenso questionamento sobre o funcionamento do cérebro. Qual seria a função do corpo caloso? Seria ele responsável pela integração das operações dos dois hemisférios? A execução da calosotomia implicaria em detecção de mais de um centro de consciência? Até que ponto as metades cerebrais seriam independentes quando fossem separadas? Poderiam elas ter pensamentos e emoções separadamente?

Diversos trabalhos com animais foram desenvolvidos por Sperry e seus colaboradores no sentido de elucidar essas indagações. A contribuição dessas pesquisas foi sig-

---

17. Esta seção foi elaborada pelo psiquiatra Marcos Romano Bicalho, a partir de estudo desenvolvido na sua dissertação de mestrado, sob minha orientação, na Universidade Federal de São Carlos. Cf. Bicalho (1997).

nificativa para o conhecimento das fibras de associação que formam o corpo caloso, assim como suas implicações teóricas. Porém, os trabalhos da equipe de Sperry não pararam na experimentação animal e prosseguiram nas investigações com humanos.

Desde o final da década de 1930, havia alguns relatos de neurocirurgiões de que a secção do corpo caloso não trazia grandes repercussões funcionais na vida das pessoas operadas. As cirurgias eram realizadas em casos graves de epilepsia. O corte do corpo caloso serviria para evitar que as descargas iniciadas em um hemisfério se generalizassem para o outro. Na época, não foram bem-sucedidas as tentativas de evidenciar déficits funcionais nesses pacientes, permanecendo um mistério as funções do corpo caloso.

Os estudos de Akelaitis[18] foram pioneiros na investigação das repercussões da calosotomia no ser humano. Achados contendo poucas manifestações comportamentais deficitárias têm sido corroborados amplamente por investigações posteriores e continuam a justificar a calosotomia como opção terapêutica para os casos intratáveis de epilepsia.

Nesse terreno parcialmente conhecido até então, Sperry, sua equipe e outros pesquisadores fizeram estudos que possibilitavam tornar esse grupo de fibras o mais conhecido dentre todos os sistemas centrais de associação do cérebro. As indagações passaram dos aspectos funcionais para as questões pertinentes às técnicas de investigar a organização e funcionamento do cérebro, abrindo novos campos de exploração da atividade cerebral.

Em 1961, Michael S. Gazzaniga, então assistente e colaborador de R. Sperry no California Institute of Technology, iniciou uma série de pesquisas que se tornariam clássicas ao realizar uma série de operações. De 11 pacientes operados, 4 foram examinados com maiores detalhes por um período prolongado de tempo. As técnicas consistiram essencialmente em explorar elementos da chamada *split-brain syndrome* de modo a testar separadamente o desempenho de cada hemisfério cerebral. A ideia central que estava por trás das experiências era poder responder de que maneira a separação dos hemisférios afetava as capacidades mentais do cérebro humano.

Para realizar esse tipo de estudo, duas abordagens foram utilizadas. A primeira consistiu em um teste a partir da percepção visual. Uma figura ou uma informação escrita era apresentada através de projeções em uma tela à direita ou à esquerda da linha média do campo visual por um décimo de segundo, tempo suficiente para a percepção da imagem, mas não o suficiente para haver movimentos oculares compensatórios capazes de trazer o estímulo para a outra metade do campo visual e, por conseguinte, ao hemisfério cerebral oposto. O paciente, sentado, fixava a visão em um ponto médio à sua frente, determinado pelos pesquisadores. É preciso lembrar que o quiasma óptico era preservado nas cirurgias. Daí essa necessidade técnica para segregar satisfatoria-

18. Apud Sperry (1968).

mente as informações que ora eram transmitidas ao hemisfério direito, ora ao esquerdo. O outro tipo de teste foi direcionado para a estimulação tátil, visto que a segregação sensitiva das mãos é considerada eficiente para a experimentação. Um objeto era colocado fora do alcance visual na mão esquerda ou direita do paciente, com o objetivo, novamente, de levar a informação a somente um dos hemisférios. A situação do experimento pode ser vista nas figuras 2.5 e 2.6.

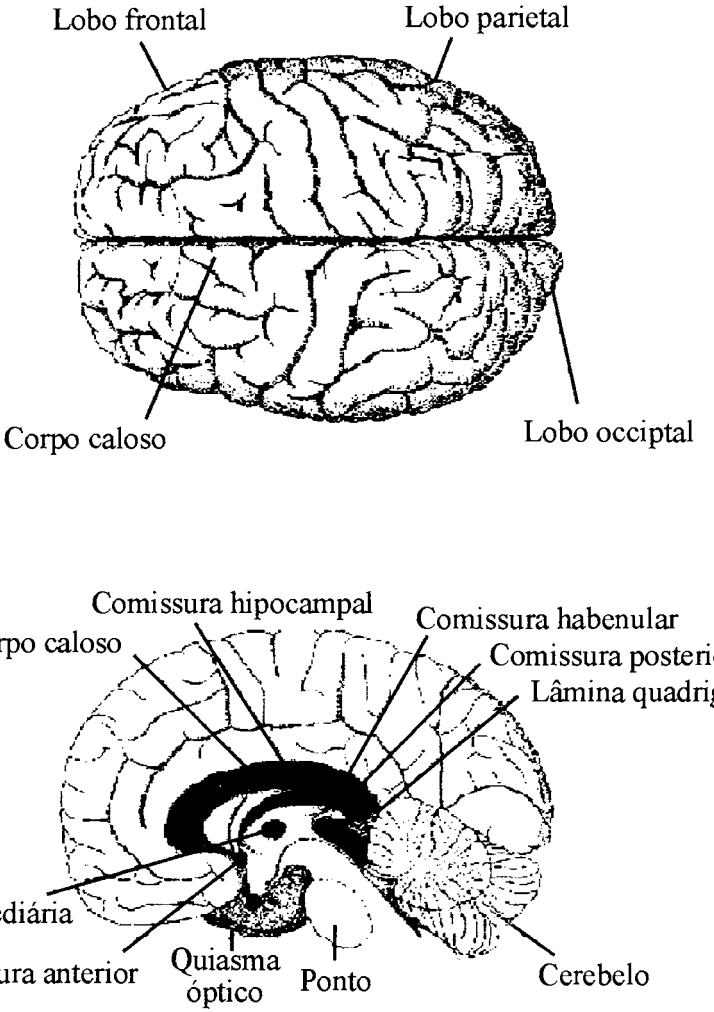

Fig. 2.2 – Corpo Caloso e as outras comissuras que conectam os hemisférios. O desenho superior mostra os hemisférios, com a posição do copo caloso em marrom. O desenho inferior mostra o hemisfério direito a partir de uma secção na linha média encefálica.

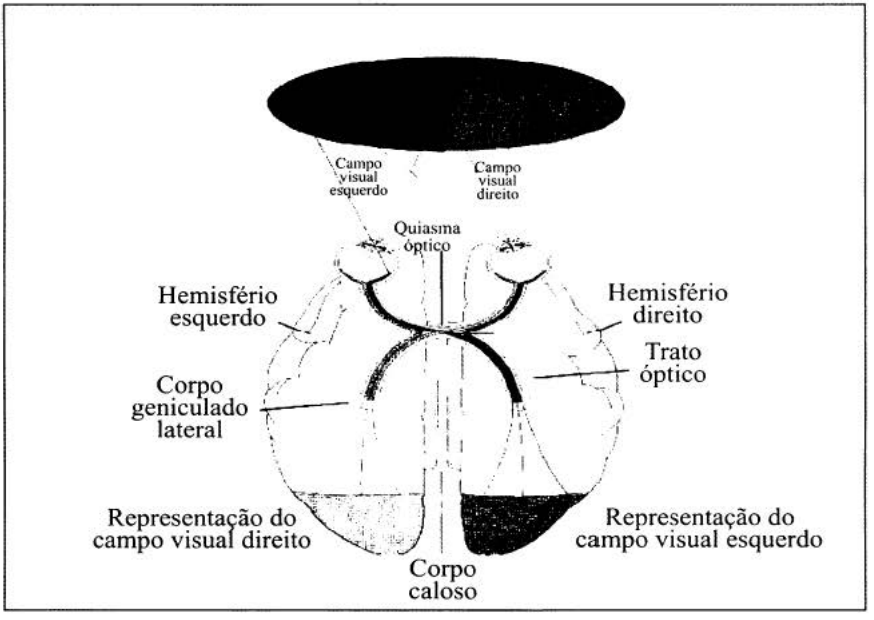

Fig. 2.3 – Campos visuais e suas relações com o sistema cerebral. O corte do corpo cialoso elimina a sobreposição dos campos visuais nas regiões corticas (Adaptado de BICALHO, 1997).

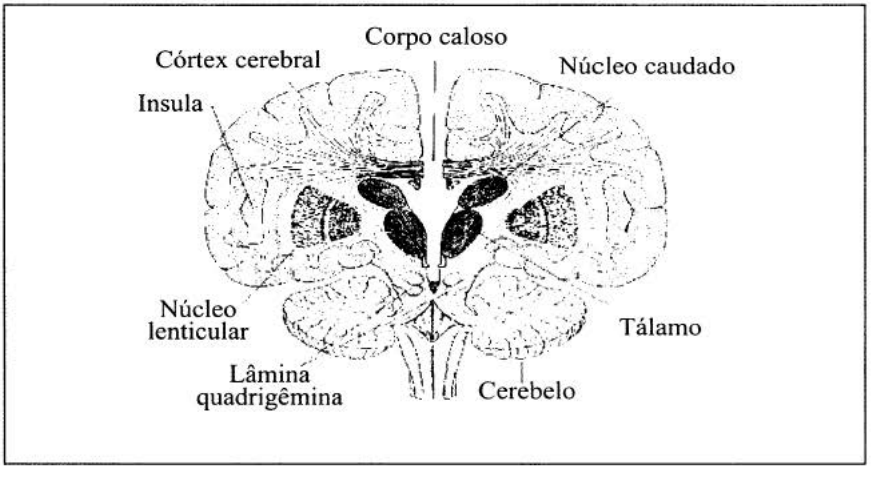

Fig. 2.4 – Grau de separação entre os centros cerebrais, destacando a secção do corpo caloso em corte frontal do cérebro (Adaptado de BICALHO, 1997).

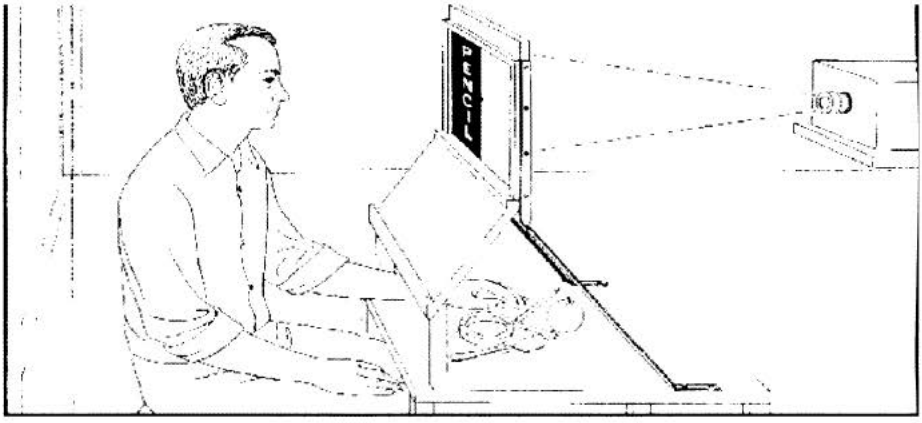

Fig. 2.5 – A resposta ao estímulo visual é testada através de "flashes" de palavras ou figuras de objetos em uma tela transparente. O examinador primeiramente checa se o indivíduo mantém o olhar fixo no ponto marcado na linha média. O examinador pode solicitar uma resposta verbal ou não verbal (pegar objetos). Os objetos estão fora do alcance visual e só podem ser identificados pelo tato (Adaptado de BICALHO, 1997).

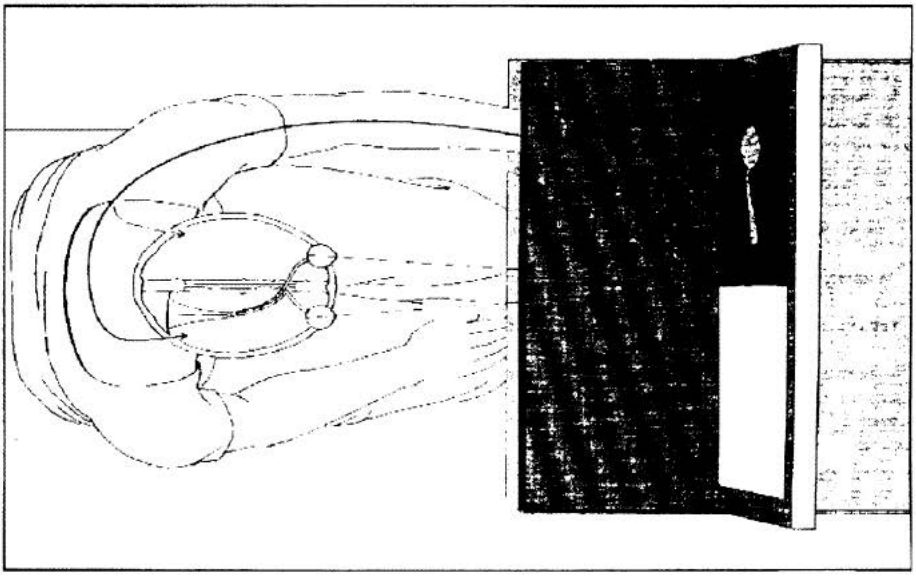

Fig. 2.6 – A associação tátil-visual é realizada pelo paciente. A figura de uma colher é apresentada ao hemisfério direito; com a mão esquerda ele seleciona uma colher detrás do aparato. A informação tátil da mão esquerda (vermelho) projeta-se para o hemisfério direito, mas um fraco sinal ipsilateral vai ao hemisfério esquerdo. Isto não é suficiente para que ele verbalize o que pegou (Adaptado de BICALHO, 1997).

As imagens de objetos ou palavras apresentadas ao hemisfério esquerdo, através do campo visual direito, eram descritas normalmente tanto pela fala quanto pela escrita. Cálculos matemáticos elementares também foram executados sem maiores dificuldades. Assim, quando a figura de uma colher era apresentada, os pacientes não só falavam o que tinham visto, como eram capazes, com a mão direita, de selecionar o objeto dentre outros que estavam fora do alcance visual. Quando se solicitava que pacientes dissessem o nome de vários objetos colocados na mão direita, eles o faziam sem dificuldade. Enfim, na estimulação do hemisfério esquerdo, as respostas eram precisas, expressas prontamente, e havia adequada manipulação de objetos e da escrita, desde que a mão direita fosse envolvida. Todos os pacientes do experimento eram destros, indicando a dominância linguística do hemisfério esquerdo.

O mesmo não ocorria quando o hemisfério direito era testado. Para as mesmas informações apresentadas ao hemisfério direito, não havia respostas faladas nem escritas. Uma figura transmitida a esse hemisfério evocava uma tentativa fortuita de responder ou mesmo nenhuma resposta. De modo análogo, um lápis colocado na mão esquerda (fora do alcance visual) poderia ser denominado isqueiro ou abridor de latas e alguns nem sequer tentaram descrevê-lo. As tentativas, muito provavelmente, não vinham do hemisfério direito, mas do esquerdo, que não tinha a percepção dos objetos. Não obstante, através de informações indiretas, arriscavam identificá-los. Alguns pacientes, após verem em seu campo visual esquerdo uma colher, eram capazes de selecionar corretamente o objeto dentre outros. Além disso, também faziam associações. Quando a figura de um cigarro era mostrada, eles acertavam, selecionando um cinzeiro em um grupo de 10 objetos que não incluía um cigarro. Entretanto, mesmo tendo acertado a escolha do objeto, segurando-o na mão esquerda, não foram capazes de dizer o nome ou de descrever o objeto.

Outro experimento particularmente interessante refere-se à apresentação da palavra "HEART" no centro do campo visual, de modo a ter "HE" somente à esquerda e "ART" à direita. Quando se perguntava o que tinham visto, respondiam "ART" (projetado no hemisfério esquerdo, responsável pela fala). Quando solicitados a apontar com a mão esquerda um dos cartões onde estava escrito "HE" e "ART", separadamente, invariavelmente apontavam "HE". Evidenciou-se que os dois hemisférios haviam visto as partes da palavra e que, dada uma oportunidade, o hemisfério direito era capaz de demonstrar suas capacidades linguísticas.

Discutindo essas respostas posteriormente, o paciente não se lembrava de ter apontado o cartão com a mão esquerda. O hemisfério esquerdo (chamado "dominante") aparentemente ignorava o que acontecia no seu par.

Esses experimentos – e vários outros semelhantes – levaram Sperry e Gazzaniga a especular que quando o cérebro é biseccionado, passamos a observar dois "eus" separados, como se o organismo, quando dividido, passasse a ter duas unidades mentais, cada uma com suas memórias e seu próprio arbítrio. Elas estariam "competindo" pelo controle do organismo.

Essa afirmação abriu campo para implicações filosóficas com respeito à independência de arbítrio das partes, sobre hierarquia do sistema orgânico e, no limite, sobre a existência de "duas pessoas" em cada um de nós.

Até que ponto poderíamos cogitar em conceber "duas pessoas" em nós? Para tanto, precisaríamos ter, pelas experiências, evidências de dois centros de consciência que pudessem interagir com o meio, questionar, informar, argumentar e assim por diante. Cada parte não poderia ser apenas receptor de informações sensitivas e gerador de movimentos corporais, mas também um centro decisório autônomo capaz de avaliar e julgar.

Em 1981, Donald MacKay, também nos laboratórios de Sperry, procurou verificar o grau de independência entre os sistemas da esquerda e da direita. O paciente J.W., de 27 anos, havia feito uma calosotomia há quatro anos. As séries de testes podem ser resumidas em uma situação onde se tentou estabelecer um diálogo entre os hemisférios.

Eis o diálogo: cada hemisfério, com sua respectiva forma de expressão (o esquerdo, oralmente; o direito, com a mão esquerda), foi treinado para participar de um jogo de adivinhação de números de O a 9 com um examinador. Para que um lado orientasse o outro nas tentativas de acertar, os sinais "para cima", "para baixo" e "OK" ficavam disponíveis. O examinador então apresentou números de um único dígito ao hemisfério direito (no campo visual direito foram colocadas letras apenas como complemento visual neutro), sendo que J.W. deveria fazer os dois papéis. Oralmente, tentaria acertar o número e com a mão esquerda apontaria os sinais para orientar o processo. Muito embora J.W. conseguisse preencher a dualidade de papéis, ele sempre se dirigia ao examinador que precisava lembrar o uso da mão esquerda para dar a resposta. Um episódio veio a mostrar que J.W. não estava apenas encenando. J.W. arriscou dizer "1" e sua mão esquerda apontou "para baixo". J.W. reclamou que não havia números mais abaixo e teve de ser lembrado da possibilidade do zero, que era a resposta correta. Um comentário espontâneo de J.W. acerca dessa situação foi particularmente esclarecedor: "Vocês estão tentando fazer de mim duas pessoas?" Ora, será que isto não estaria revelando que temos dois fluxos (ou mais?) de consciência que são integrados pelas comissuras?

As experiências com os *split-brain patients* abriram um campo de exploração novo para os filósofos da mente. Acreditou-se que, pela primeira vez, a filosofia da mente poderia utilizar-se de dados empíricos para resolver um de seus problemas fundamentais, qual seja, o problema da relação entre mente e corpo. Pois, se dividirmos cérebros e, após essa divisão, encontrarmos também uma divisão na mente, poderemos então refutar o argumento cartesiano da assimetria entre o físico e o mental baseado na indivisibilidade desse último. Contudo, essa estratégia esbarra numa dificuldade: não lidamos apenas com dados empíricos, mas também com a interpretação desses últimos; uma interpretação que força a opção por um ponto de vista ou por outro, mas não nos permite tirar uma conclusão definitiva.

Não discutiremos aqui os estudos que fazem inferências sobre as habilidades especiais de cada hemisfério cerebral e a dominância de um sobre o outro no controle de

certos comportamentos humanos. Não focaremos, portanto, o tópico sobre que tipos de mentes surgem nas experiências. Estamos preocupados com o número de mentes que esses pacientes têm. Discutiremos até que ponto experiências com os *split-brain patients* nos permitem confirmar a proposição de que cada paciente tem uma mente e é uma pessoa, muito embora, em algumas ocasiões, sua consciência esteja dividida. Isto implica em não somente questionar se "seria possível dividir o mental com a separação dos hemisférios cerebrais", mas também "que critérios devemos usar para caracterizar a unidade da consciência".

Iniciemos nossa reflexão recordando a proposição cartesiana acerca da indivisibilidade do mental que aparece no parágrafo 33 da *Meditação VI*.

> [...] há grande diferença entre espírito e corpo, pelo fato de ser o corpo, por sua própria natureza, sempre divisível e o espírito inteiramente indivisível. Pois, com efeito, quando considero meu espírito, isto é, eu mesmo, na medida em que sou apenas uma coisa que pensa, não posso aí distinguir partes algumas, mas me concebo como uma coisa única e inteira[19].

O filósofo Thomas Nagel defendeu uma perspectiva cética em relação à utilização dos experimentos com calosotomia, para nos ajudar a esclarecer a tese cartesiana da assimetria entre o físico e o mental. Em seu artigo "Brain Bisection and the Unity of Consciousness" [Bisecção cerebral e a unidade da consciência] (1971) ele argumentou que diante das pesquisas com comissurotomia nosso conceito de *unidade da pessoa* torna-se confuso. Para substanciar esse ponto de vista, Nagel argumenta que o fluxo de consciência altera-se com as divisões cerebrais e que temos dificuldade em assumir a possibilidade de mais de uma mente em nós mesmos porque utilizamos o conceito primitivo e essencialmente intuitivo da unidade da pessoa, que não poderia ser desafiado por esses experimentos de secção cerebral. A noção de uma unidade da pessoa, por ser primitiva, persistiria *apesar* dos possíveis resultados revelados pelas comissurotomias.

Nesse sentido, Nagel sustenta que experimentos com pacientes comissurotomizados não seriam um caminho para decidirmos nem a favor nem contra a assimetria proposta por Descartes. Os experimentos com comissurotomias revelam que podemos ter pelo menos duas mentes, ou pelo menos dois hemisférios, um que seria capaz de "falar" e outro não. Se esses dois hemisférios ou essas duas mentes funcionam em paralelo ou se eles se comunicam para formar apenas uma mente não seria relevante para o problema com o qual nos defrontamos. O que conta, em qualquer dessas hipóteses, seria o fato de o sistema, sendo ele uno ou múltiplo, falar de si mesmo usando a primeira pessoa do singular, ou seja, referindo-se a si mesmo como sendo uma unidade. Concebemos a nós próprios como sendo uma unidade, mesmo que isto seja uma ficção – uma falsa representação de nós mesmos que tende a persistir, apesar de não encontrar apoio nos

19. Cf. *Meditações*, p. 139 da edição brasileira.

dados experimentais revelados pela neurociência. O verdadeiro problema *filosófico* que teríamos de enfrentar seria, então, saber por que essa ficção ou ideia primitiva de unidade da pessoa persiste, apesar de ser, muito provavelmente, fictícia. Dessa perspectiva, dados empíricos seriam pouco importantes para decidir se a assimetria proposta por Descartes seria solapada pela investigação neurofisiológica. Viveríamos, sobreviveríamos e nos comunicaríamos uns com os outros baseados nesse tipo de ficção útil, que aparece sempre que abrimos a boca para falarmos de nós mesmos como sendo uma unidade e não uma pluralidade de sistemas neuronais.

A ideia de uma unidade do mental parece estar correta, mas acreditamos que sua base pode ser explicada neurocientificamente, sem ter de recorrer a ficções, unidades da pessoa tomadas como intuições primitivas ou outros artifícios filosóficos. Se o mental não é divisível ou se a unidade da pessoa não é divisível, não podemos saltar disto para a conclusão de que ambos são imateriais sem antes percorrer outras alternativas que nos são fornecidas pela neurociência. Nagel ignorou essas alternativas, ao pressupor um modelo de cérebro baseado na ideia de que haveria duplicação das bases físicas correspondentes às funções mentais nos dois hemisférios e que entre esses ocorreria sempre uma cooperação – a chamada cooperação inter-hemisférica. De acordo com essa perspectiva, a cada região do cérebro corresponderia uma atividade mental: esse é o chamado *localizacionismo*, por oposição a um modelo no qual todas as partes do cérebro entrariam, de tempos em tempos, numa espécie de mutirão para produzir atividades mentais dos mais variados tipos (*integracionismo*).

Iniciemos fazendo um experimento mental simples sugerido por C.E. Marks, no seu livro clássico *Comissurotomy, Consciousness and the Unity of Mind* [Comissurotomia, consciência e a unidade da mente] (1981). Existem técnicas, desenvolvidas por Wada[20], para anestesiar seletivamente cada hemisfério. Apesar de não ser possível fazê-lo atualmente com as comissuras, poderíamos perfeitamente conceber a possibilidade de realizar esse procedimento. Supondo que nossas comissuras fossem anestesiadas por dez minutos, teríamos o mesmo comportamento dos pacientes comissurotomizados durante esse intervalo. Teríamos as mesmas evidências da divisão da consciência, implicando em uma separação dos fenômenos mentais análoga à dos *split-brain patients*, assim como a mesma coerência interna de cada hemisfério. Mas não seria isto o que ocorreria: a consciência desunida, por dez minutos, que simularia a ausência funcional do caloso não ocasionaria a dicotomia do mental. A interrupção momentânea da comunicação entre os dois hemisférios não revelaria a duplicidade de fluxo de consciência. Some-se a isto o fato de que não se verifica dicotomia do mental nos indivíduos que nascem sem o corpo caloso.

Quando as comissuras principais são seccionadas, verificamos uma síndrome caracterizada pela quebra da unidade da consciência somente nas circunstâncias em que

20. Apud Marks (1981).

as informações são seletivamente enviadas a cada hemisfério. Funções específicas, relacionadas a áreas sensitivas, quando segregadas, apresentam respostas independentes, mas isto não implica na demonstração de dois centros decisórios autônomos. Os trabalhos com J.W. são bastante ilustrativos nesse sentido.

J.W. foi o paciente citado por MacKay por ter reunido as melhores condições experimentais. Os outros pacientes, conforme apontou Myers, apresentavam uma série de problemas que prejudicavam a uniformidade de controle experimental. É preciso lembrar que a extensão das anormalidades apresentadas variaram conforme o tipo de cirurgia executada. As dificuldades de homogeneidade experimental foram encontradas desde as lesões cerebrais prévias, diferentes em vários pacientes, até incertezas do alcance cirúrgico de cada procedimento. Todos os pacientes avaliados, conforme citamos anteriormente, eram destros e, portanto, de acordo com as teorias predominantes naquele momento, foram relacionados com a dominância linguística à esquerda.

Gazzaniga entendia haver uma grande diferença entre os hemisférios com relação às capacidades linguísticas, sendo o hemisfério direito praticamente nulo nessa área. Concluiu assim:

> Realmente, poder-se-ia argumentar que as habilidades cognitivas de um hemisfério direito desconectado e sem linguagem são vastamente inferiores às habilidades cognitivas de um chimpanzé[21].

À luz das insuficiências experimentais, Myers lançou críticas em relação às inferências feitas por Gazzaniga de que o hemisfério direito seria inferior ao esquerdo em várias funções, especialmente na linguagem. Suas observações receberam confirmação experimental em um recente trabalho de Lutsep (1995) de que o próprio Gazzaniga participou, mostrando que a potencialidade linguística do hemisfério direito, dentre outras, era equivalente em uma paciente com esse hemisfério dominante e calosotomia. Foi a primeira pessoa operada (1988) na condição invertida em relação aos outros pacientes, indicando que o sistema nervoso central segue um desenvolvimento próprio, hierarquizado, mas não especificamente predeterminado, no que diz respeito à localização de certas funções.

Não podemos perder de vista que os testes posicionaram os pacientes de tal modo que as informações a eles emitidas eram altamente segregadas, sensibilizando partes diferentes do sistema. Esse aspecto é relevante, uma vez que se pode ter a falsa impressão de que centros autônomos são integrados para se apresentarem como um *só*. Em nenhum momento dos experimentos observaram-se centros decisórios, com autorreferência, distintos. Não se evidenciou a constituição de conflitos característicos de entidades mentais distintas. Se houvesse uma correlação exata entre fluxos de consciên-

21. Gazzaniga (1983, p. 536).

cia individuais e atividade cortical unilateral, a separação dos hemisférios deveria proporcionar o aparecimento de dois fluxos de consciência, o que não ocorreu.

Entretanto, vale notar que nenhum dos hemisférios teve consciência do conteúdo das informações do outro, como teve de seu próprio. Pertinente é a observação de Marks[22]:

> O fato crucial é que nossas explicações ordinárias falham somente quando as diferentes aderências são fornecidas a cada hemisfério. A aderência é de tal tipo que pode ser recebida pelos dois hemisférios e produzir diferentes respostas e nenhuma integração é possível por outras vias que não as comissuras. Essas condições são altamente artificiais; um meio ambiente normal não as fornece.

Há ainda um outro aspecto a se questionar: se uma mente, para ser una em seu fluxo de consciência, precisa necessariamente ter acesso imediato a todos os seus estados de consciência.

Situações do cotidiano nos respondem. Imaginemos um indivíduo falando ao telefone enquanto uma série de outras coisas acontecem à sua volta. Ou um cidadão pedalando uma bicicleta pelas ruas, lendo cartazes, olhando os carros, entre outras coisas. Será que essas pessoas estarão conscientes nesses momentos de todos os seus estados internos quando os estímulos são variados e simultâneos? Assim sendo, podemos sustentar uma resposta negativa para o problema que enunciamos no parágrafo anterior.

Podemos inferir legitimamente, baseados nos fatos e nos estudos de neurofisiologia, que existem graus diferentes de consciência. Os tipos diferentes de aferência no sistema nervoso central repercutem distintamente no que concerne à consciência. É notório que à visão o sistema reserva uma área cortical maior que para os outros sentidos. Somos mais "impressionados" (do ponto de vista fisiológico) pelas cores do que pelos sons. A consciência de um aroma é diferente daquela em relação aos movimentos das alças intestinais etc.

O localizacionismo parece ser o grande pressuposto do qual Nagel deriva todos os seus argumentos. Nagel interpretou os resultados como indicadores de que cada hemisfério possui um alto grau de independência um do outro. Em vários aspectos, isto não deixa de ser verdadeiro. Sua interpretação apontou para uma cooperação entre as metades através de uma constante intercomunicação. Contudo, o sentido que Nagel deu à palavra cooperação não é adequado. Ele a utilizou tratando os hemisférios como se fossem dois amigos que, apesar de distintos e independentes, "cooperam" entre si para viverem bem na mesma casa. Entendemos que eles cooperam entre si porque são um órgão integrado.

Essas experiências mostram que o sistema nervoso central é especializado. Existem regiões corticais diferentes para recepção de estímulos sensitivos diferentes.

---

22. Marks, op. cit., p. 39.

Mostrou-se a dominância hemisférica, seja à esquerda ou à direita, para determinadas funções. Ficou patente que o corpo caloso não é um órgão inerte. Sem ele, as pessoas vivem um cotidiano aparentemente inalterado, mas não podemos esquecer que existem as outras comissuras que exercem um papel semelhante e que buscam a integração, compensando parcialmente a ruptura do caloso e são da mesma natureza das fibras do caloso. O fato é que por todos os experimentos permeia uma sensível vocação do sistema integrar-se. Trata-se de uma característica essencial, marca da unidade.

Essa unidade não precisa pressupor a imaterialidade do mental como queria Descartes: sua natureza pode ser explicada em termos neurofisiológicos, sem ter de recorrer a ficções filosóficas ou ideias primitivas e intuitivamente dadas de uma "unidade da pessoa". Resta saber, contudo, se a pesquisa neurocientífica poderia nos fornecer uma resposta conclusiva em favor ou contra a assimetria proposta por Descartes, ou seja, saber se da divisão do cérebro poderíamos, em circunstâncias experimentais, derivar a divisão do mental como sua consequência. Estaríamos, assim, dando um passo final, decisivo, para tornar o problema das relações entre mente e cérebro um problema científico, ou seja, empírico e não mais filosófico. Teríamos encontrado uma simetria inicial, mas fundamental para sustentar que mente e cérebro seriam a mesma coisa.

Contudo, as circunstâncias experimentais das quais podemos derivar conclusões baseadas estritamente na investigação neurocientífica podem ser desfavoráveis e, em última análise, impedir-nos de confirmar a hipótese da existência dessa simetria, de onde se derivaria, como passo seguinte, a identidade entre mente e cérebro. As experiências de que dispomos, com pacientes comissurotomizados, mostram-nos que podemos descrevê-los não como sendo "duas pessoas" ou dois "centros decisórios independentes", mas como pessoas que, por ter uma parte de seu corpo seccionada, são capazes, em certas circunstâncias, de exibir formas particulares de ausência de consciência perceptiva. Isso porque o corte do caloso traz à tona que determinadas funções corticais são executadas por áreas diferentes e que os hemisférios não são um a imagem especular do outro. Mas isto nos permite inferir, ademais, que a integração do sistema supervisor faz-se por um intenso tráfico de informações que envolve, além das estruturas corticais, outras como o sistema límbico/hipotalâmico e núcleos subcorticais, os quais não são separados pela calosotomia. Seriam as comissurotomias realmente capazes de dividir o sistema supervisor/integrador que proporciona as experiências conscientes? Seria esse um problema específico desse tipo de cirurgia ou será que poderíamos questionar a possibilidade de existência de *qualquer tipo* de cirurgia que pudesse, em última análise, dividir o sistema supervisor/integrador? Nesse caso, poderíamos questionar até que ponto o seccionamento de partes do cérebro seria uma estratégia adequada da qual poderíamos inferir possíveis seccionamentos da experiência consciente. A grande dificuldade enfrentada por esse tipo de pesquisa está no fato de que, ao tentarmos correlacionar elementos da experiência consciente, ou sua ausência, com partes danificadas ou seccionadas do cérebro, estamos interferindo num sistema tão ricamente integrado que esse procedimento pode nos levar a inferências errôneas, inconclusivas ou até mesmo impossibilitar a realização de experimentos cruciais para

averiguarmos nossas hipóteses. Esse é um grande desafio que a neurociência vem tentando contornar pelo desenvolvimento de técnicas de neuroimagem, abrindo o campo para o aparecimento e a consolidação progressiva da *neurociência cognitiva* da qual falaremos mais adiante neste livro. Essas dificuldades experimentais que ainda enfrentamos não significam, contudo, que as comissurotomias tenham deixado de ser um rico material para exploração filosófica e metodológica.

O QUE LER

O volume sobre Descartes da coleção "Os Pensadores", da Abril Cultural, reúne as principais obras desse pensador, em excelente tradução para a língua portuguesa.

Sobre a teoria dos autômatos em Descartes:

GUNDERSON, K. *Mentality and Machines*

Sobre comissurotomia:

MARKS, C.E. *Comissurotomy, Consciousness and the Unity of Mind*

CAPÍTULO III

MATERIALISMO E TEORIAS DA IDENTIDADE

Examinaremos neste capítulo o *materialismo* ou *fisicalismo*. Esse se tornou, a partir dos anos de 1950, uma das correntes mais fortes da filosofia da mente no século XX. Sua inspiração é a perspectiva crescente de que novas descobertas no campo da neurociência permitir-nos-ão, mais cedo ou mais tarde, concluir que a mente é algum tipo de manifestação da atividade do cérebro.

O materialismo nos repugna pela sua crueza. Toda nossa vida mental nada mais seria do que uma grande variação dos estados químicos e físicos de nosso cérebro. Nossas angústias, desejos e intenções seriam apenas um produto do cérebro e supor que tenham existência autônoma não passaria de uma ilusão. Poderíamos também interferir nesses processos químicos usando drogas e, nesse caso, depressões ou melancolias deixariam de ser crises existenciais ou profundos conflitos de valores para se tornarem apenas desequilíbrios orgânicos passageiros, que poderiam ser curados da mesma maneira que nos restabelecemos de uma diarreia tomando alguns comprimidos.

Mas ao supor que *todos* os nossos estados mentais, inclusive aqueles correspondentes a "sustentar a veracidade de uma teoria materialista da mente", nada mais seriam do que estados cerebrais, não estaríamos caminhando para um paradoxo? Pois ao sustentar que o estado mental correspondente à proposição "Todos os estados mentais são estados cerebrais" nada mais é do que um estado cerebral entre outros, não estaríamos correndo o risco de afirmar que a veracidade dessa proposição seria apenas um estado cerebral efêmero que poderia mudar no instante seguinte – ou ser

alterado pela ingestão de algum tipo de remédio? E, nesse sentido, o materialismo não seria uma tese autocontraditória?[1]

O materialismo comporta algumas variações teóricas que delineiam possíveis soluções para o problema mente-cérebro: estados mentais *são* estados cerebrais (teorias da identidade) ou *são redutíveis* a estados cerebrais (reducionismo) ou eles *emergem* de estados cerebrais (emergentismo ou teorias da superveniência). Ao percorrer esse tipo de literatura, o leitor deve também ficar atento para o fato de que frequentemente a palavra monismo é utilizada no lugar de materialismo, significando, na maioria das vezes, *monismo materialista*, ou seja, a doutrina que sustenta que estados subjetivos nada mais são do que um tipo de manifestação do mundo físico. A palavra *fisicalismo* é também frequentemente empregada como sinônimo de materialismo, na acepção de que estados mentais são um produto da atividade física do cérebro. A afirmação popular de que a mente seria algum tipo de energia constitui uma variação do materialismo (ou fisicalismo), pois a energia é uma entidade física, a não ser que entendamos por energia algum tipo de *élan vital* ou quintessência, ao modo cartesiano.

Algumas distinções técnicas e terminológicas

Antes de iniciarmos nossa apresentação e discussão das variedades do materialismo contemporâneo, precisamos introduzir algumas distinções técnicas e terminológicas que são usadas com muita frequência na filosofia da mente. Sem elas não poderemos, mais adiante, entender a verdadeira envergadura do projeto materialista.

Quando consideramos a vida mental de um organismo, nós a descrevemos como um conjunto de *estados mentais* e explicamos seu comportamento a partir desses estados. De forma geral, estados mentais podem ser classificados em dois tipos: *estados qualitativos* ou *qualia* e *atitudes proposicionais*. Os estados qualitativos (*qualia*) correspondem, em geral, a sensações de algum tipo. As atitudes proposicionais correspondem a crenças, desejos, medos, dúvidas etc. São chamadas assim porque envolvem a ideia de existir um agente ou sujeito que tem uma determinada atitude em relação a uma certa proposição. Consideremos, por exemplo a seguinte proposição: "Há um besouro no quintal". Posso ter diferentes atitudes em relação a essa proposição: posso *acreditar* que há um besouro no quintal, posso *duvidar* que haja um besouro no quintal, posso *temer* que haja um besouro no quintal e assim por diante. Acreditar, duvidar, temer etc., caracterizam diferentes *atitudes* em relação à proposição "Há um besouro no quintal".

Uma maneira simplificada de dizer o que é uma proposição é caracterizá-la como sendo um pensamento. Proposições (ou pensamentos) são, frequentemente, expressos através de sentenças da linguagem, mas há uma diferença entre proposição e sentença.

1. Um argumento similar foi formulado, pela primeira vez, por Haldane (1932).

"Está chovendo" e *It rains* são sentenças diferentes, embora expressem a mesma proposição, ou seja, o pensamento de que "está chovendo". Embora os filósofos nunca tenham chegado a um consenso sobre o que seja o pensamento e muito menos sobre o que seja uma representação, podemos afirmar que, de modo geral, proposições são representações, ou seja, conteúdos mentais acerca de alguma coisa, mesmo que essa possa não existir concretamente. Nesse sentido, atitudes proposicionais e proposições são dotadas de intencionalidade ou direcionalidade de que falamos no capítulo anterior[2].

O estudo do comportamento das proposições e de seu aspecto representacional foi uma preocupação constante da filosofia da linguagem e da filosofia da mente no século XX. Esse estudo revelou que atitudes proposicionais podem gerar os chamados *contextos opacos* ou *contextos intensionais* (intensionais com s). Consideremos as seguintes sentenças, a partir da peça *Édipo-Rei* de Sófocles[3]:

A: *Édipo casa-se com Jocasta.*

B: *Édipo deseja que Jocasta se torne sua esposa.*

No caso da sentença A podemos substituir o termo "Jocasta" pelo termo "a mãe de Édipo" e o valor de verdade da sentença (isto é, se ela é verdadeira ou falsa) continuará sendo o mesmo. Nesse caso temos:

A' – *Édipo casa-se com a mãe de Édipo.*

O valor de verdade de A e de A' é o mesmo.

Tomemos agora a sentença B e substituamos o termo "Jocasta" pelo termo "a mãe de Édipo". Nesse caso, teremos a sentença:

B' – *Édipo deseja que a mãe de Édipo se torne sua esposa.*

B e B' não têm o mesmo valor de verdade. Aliás, é precisamente por isso que a vida de Édipo tornou-se trágica. "Jocasta" e "a mãe de Édipo" são termos que designam a mesma coisa no mundo, mas o valor de verdade é alterado: "Jocasta" e "a mãe de Édipo" são modos diferentes de *representar* uma mesma coisa no mundo. A representação alterou o valor de verdade de B'. Édipo jamais poderia admitir que desejava casar-se com sua mãe, mas admitiria que desejava casar-se com Jocasta – embora "Jocasta" e "mãe de Édipo" designem a mesma coisa no mundo. Quando ocorre uma situação desse tipo, temos um *contexto opaco* ou um *contexto intensional* (com s). Aliás, se Édipo soubesse que "Jocasta" e a "mãe de Édipo" designavam a mesma coisa, não teria havido tragédia e ele não teria tido que arrancar seus olhos.

O que isto tem a ver com o problema das relações entre mente e cérebro? Consideremos mais um exemplo. Imaginemos uma situação na qual não se soubesse que um raio no céu era a mesma coisa que uma descarga elétrica. Nossas crenças acerca

2. Cf. o final da seção "A filosofia pós-cartesiana: uma história abreviada".

3. Exemplo inspirado em Hannan (1994).

de raios e de descargas elétricas seriam diferentes; o que se poderia dizer acerca de raios não se aplicaria às descargas elétricas. Acreditaríamos estar diante de fenômenos diferentes; estaríamos diante de um contexto opaco ou intensional (com s). Mas, quando se descobriu que raios e descargas elétricas eram a mesma coisa, o contexto opaco se desfez.

Ora, o mesmo poderíamos esperar da descrição ou representação da mente e do cérebro: ao descrever fenômenos mentais e fenômenos cerebrais estaríamos diante de um contexto opaco, de um modo diferente de descrever o mesmo tipo de fenômeno. Mas, na medida em que a ciência avança poderíamos desfazer o contexto opaco, ou seja, progressivamente verificar que mente e cérebro são a mesma coisa, do mesmo modo que um raio no céu e uma descarga elétrica. Fenômenos mentais seriam apenas um modo de representar fenômenos cerebrais. Essa é a grande esperança, mas, ao mesmo tempo, o grande calcanhar de Aquiles do materialismo contemporâneo: desfazer um suposto contexto intensional, mostrando que mente e cérebro são um mesmo objeto descrito de modos diferentes. Para usar uma linguagem mais técnica: que o *intensional* (a mente, suas crenças, desejos etc.) poderia ser reduzido ao *extensional* (o cérebro). O que isto significa?

A extensão de um termo é a classe das coisas às quais esse termo se refere. A extensão do termo "os presidentes americanos entre 1961 e 1968" é a classe contendo dois membros: Kennedy e Johnson. Considere agora os termos "estrela da manhã" e "estrela da tarde". Sabemos que ambos os termos, "estrela da manhã" e "estrela da tarde", designam o planeta Vênus. Vênus é a *extensão* desses termos. Mas "estrela da manhã" e "estrela da tarde" têm significados ou *intensões* diferentes. Alguém pode "acreditar que a estrela da manhã é bela" sem, entretanto, "acreditar que a estrela da tarde seja bela". As intensões afetam o valor de verdade da sentença (como já vimos no caso de B e B'), mas se usarmos "Vênus" em vez de "estrela da manhã" ou "estrela da tarde" o valor de verdade das sentenças não seria alterado. Note-se que a linguagem da ciência, sobretudo da física, é estritamente extensional. Nesse domínio o significado de um termo é univocamente dado pela sua extensão. Se houver intensões ou significados, esses precisam ser redutíveis a algo extensional, ou seja, é preciso delimitar a classe das coisas às quais o termo se refere.

Em alguns contextos, encontrar a extensão de um termo pode ser particularmente problemático. Considere, por exemplo, o termo "comportamento inteligente". Como delimitar a classe das coisas às quais o termo "inteligente" se refere? Será que à palavra "inteligência" corresponde alguma classe específica de objetos no mundo? Podemos reconhecer um comportamento como sendo inteligente, mas dificilmente conseguiríamos delimitar uma classe de objetos no mundo correspondentes ao termo "inteligência". Entretanto, o termo "inteligente" ou "inteligência" é amplamente utilizado em psicologia, ou seja, usado apesar de não ter sido reduzido, até agora, a algo extensional. A psicologia estaria entremeada de termos intensionais, o que, segundo alguns, impediria que ela se tornasse uma ciência como a matemática ou a física. A última

grande – e malsucedida – tentativa de eliminar termos intensionais da linguagem da psicologia foi feita pelo behaviorismo há algumas décadas. Para o behaviorista, a psicologia deveria se limitar ao extensional, o que a restringiu a ser uma ciência de comportamentos observáveis.

Os materialistas ou fisicalistas não são necessariamente behavioristas, embora o inverso se aplique, isto é, que os behavioristas são materialistas. Mas ambos partilharam um projeto comum, qual seja, a redução dos termos intensionais da psicologia à classe de objetos aos quais eles se referem, ou seja, a sua extensão. No caso dos materialistas e fisicalistas, esse projeto consiste em mostrar que a extensão dos estados mentais é um conjunto de estados cerebrais. Estados mentais seriam, quando muito, uma intensão ou um modo provisório de falarmos de estados cerebrais. Se há ainda um hiato entre esses dois modos de descrição ou entre essas duas representações de um mesmo objeto (mente e cérebro), ele poderá ser progressivamente suprimido pelo avanço da investigação científica.

Nesse projeto de redução ou identificação do mental ao físico estabelece-se, frequentemente, mais uma distinção: entre *type-type identity* (identidade entre tipos) e a *token-token identity* (identidade ponto a ponto). No primeiro caso busca-se uma identidade entre tipos de estados mentais e tipos de estados cerebrais: por exemplo, quando sinto tristeza, esse é um tipo de estado mental que deve corresponder a um tipo de estado cerebral correspondente à ativação de um determinado conjunto de neurônios. No segundo caso, estabelece-se que a tristeza deve corresponder a *algum* evento cerebral; ou seja, pode haver vários conjuntos de neurônios que, quando ativados, produzem em mim a sensação de tristeza. Em outras palavras, no caso da *token-token identity*, o mesmo estado mental pode ser produzido por diferentes estados cerebrais; o único requisito de identidade é que a um estado mental corresponda algum tipo de base física, seja ela qual for. Conforme veremos no capítulo VI, a *token-token identity* será defendida pela maioria dos partidários do funcionalismo ou do modelo computacional da mente.

Teorias da identidade

As teorias da identidade surgiram, independentemente, nos Estados Unidos e na Austrália durante as décadas de 1950 e 1960. Seus maiores proponentes foram H. Feigl (1958), U.T. Place (1970) e J.J.C. Smart (1962). Essas teorias propõem que estados mentais são idênticos a estados cerebrais ou estados do sistema nervoso. Em outras palavras, essa teoria poderia ser resumida através da seguinte equação:

estados mentais = estados cerebrais

U.T. Place, um dos principais proponentes da teoria da identidade, ilustra essa equação através da seguinte metáfora: suponha que você está observando uma nuvem. O que aparece a você é uma nuvem, com sua forma, sua cor branca etc. Contudo, se você refinar a sua observação através de um instrumento científico verá que a nuvem é, na verdade, composta de uma multiplicidade de gotículas de água, ou seja, ela é ape-

nas uma aparência, sua realidade é o imenso conjunto de gotículas de água. O mesmo ocorre com aquilo que chamamos de "estados mentais": eles são, na verdade, a aparência de um imenso conjunto de eventos neurais no cérebro. A terminologia psicológica e a terminologia física designam, na verdade, um único e mesmo conjunto de objetos, a saber, os eventos neurais no cérebro.

Smart observa que ao falarmos de identidade entre estados mentais e estados cerebrais estamos, na verdade, falando de uma *identidade contingente*. Esse é um tipo peculiar de identidade, ocasional e não necessária, distinta da identidade lógica que faz com que a proposição "2 + 2 = 4" seja válida em todos os mundos possíveis. Ou seja, no caso da identidade lógica, "2 + 2" ser idêntico a "4" é algo que se segue da própria definição de "2" e da definição de "+"; a identidade ocorre dentro de uma linguagem e a partir do significado que os termos dessa linguagem possuem. Isto torna essa identidade válida para todos os mundos possíveis, se mantivermos, é claro, a mesma linguagem e o mesmo significado de seus termos.

Por oposição, a identidade contingente é uma *identidade de fato*: estados mentais são idênticos a estados cerebrais porque *assim ocorre* nesse mundo em que vivemos. Smart observa que essa identidade é uma *ocorrência empírica*, algo que nada tem a ver com as propriedades da linguagem que usamos para falar de estados mentais ou de estados cerebrais. Como essa identidade decorre de um fato empírico e não do significado dos termos da linguagem que usamos, é possível que alguém, por não estar informado desse fato empírico, negue que estados mentais sejam estados cerebrais. Isto abre a possibilidade de entendermos por que um lavrador, por exemplo, negaria que seus estados mentais fossem estados cerebrais: por estar desinformado acerca desse fato, da mesma maneira que alguém pode ignorar que um relâmpago é uma descarga elétrica ou que, pelo menos nesse mundo, todos os relâmpagos são idênticos a algum tipo de descarga elétrica.

Ora, que tipo de objeção se faz a essa teoria da identidade proposta por Smart? A mais comum baseia-se na chamada Lei de Leibniz, que sustenta que, se duas coisas ou entidades são idênticas, a elas podemos atribuir exatamente as mesmas propriedades. Se a Lei de Leibniz for verdadeira – e, pelo menos intuitivamente ela parece ser – haveria pelo menos alguns tipos de assimetrias entre estados mentais e estados cerebrais que poderiam forçar-nos a abandonar a teoria da identidade.

A primeira delas consiste em mostrar que essa identidade leva-nos, inevitavelmente, a um conjunto de paradoxos semânticos. Se estados mentais ocorrem no cérebro e são idênticos a atividade cerebral, a eles seria legítimo atribuir propriedades características dos neurônios, como, por exemplo, "umidade", ou "capacidade de transmitir corrente elétrica". Ora, seria a minha tristeza úmida, seria ela capaz de transmitir corrente elétrica ou teria cabimento afirmar que ela ocorre a 5cm de distância do hemisfério esquerdo de meu cérebro?

Um outro tipo de objeção consiste em afirmar que se estados mentais são idênticos a estados cerebrais e se esses últimos são coisas físicas que ocorrem no espaço, o mes-

mo teria de ser dito acerca dos primeiros. Críticos da teoria da identidade sugeriram que se supusermos que estados mentais ocorrem no espaço, alguns paradoxos surgiriam, como, por exemplo, eu ter de supor que a sentença "minha ansiedade ocorre a cinco centímetros do hemisfério esquerdo de meu cérebro" faz sentido; ou que uma sentença semelhante a "meu sonho ocorreu a 5 cm do hemisfério esquerdo do cérebro" faria sentido. Esses paradoxos, porém, são apenas aparentes. Através de técnicas de neuroimagem, poderíamos detectar quais são as áreas cerebrais que são ativadas quando sofro de ansiedade ou quando sonho – técnicas que ainda não estavam disponíveis à época em que essas críticas foram formuladas. O que nos resta saber é se essas técnicas realmente permitem sustentar a veracidade das teorias da identidade, uma discussão que deixaremos para o último capítulo deste livro.

A assimetria mais forte entre estados mentais e estados cerebrais que não é resolvida pela teoria da identidade consiste em afirmar que os primeiros são intensionais (com s). Smart não teria podido responder a esse tipo de objeção, que implica que um mesmo estado cerebral poderia corresponder a estados mentais diferentes ou com um significado (ou intenção) diferente. Seria possível encontrar os correlatos neuronais de diferentes significados, mas não seria possível fazer a operação inversa se um mesmo estado cerebral pudesse produzir estados mentais com *intensões* diferentes. A uma diferença de intenção corresponderia uma diferença de significado que não seria escrutável em termos de uma diferença de estado cerebral. Um mesmo estímulo perceptual (resultando num estado cerebral) poderia produzir um estado mental equivalente a "Jocasta" ou a "mãe de Édipo" e, a partir da detecção desse estado cerebral, não seríamos capazes, contudo, de determinar precisamente que tipo de conteúdo mental (ou representação) estaria sendo produzido, "Jocasta" ou "mãe de Édipo".

Teorias reducionistas

Há uma diferença entre teorias da identidade e teorias reducionistas. Enquanto as primeiras afirmam que estados mentais *são* estados cerebrais, as teorias reducionistas afirmam que estados mentais *podem ser reduzidos* a estados cerebrais. Estados cerebrais poderiam ser descritos através de teorias físicas, o que torna o reducionismo um imenso programa teórico que visa reduzir teorias e termos psicológicos a teorias e termos físicos. A redução que se busca pressupõe a possibilidade de uma tradução e de uma redução entre as diversas teorias científicas, começando pela psicologia, passando pela biologia, pela química e terminando na física. Nessa perspectiva, teorias e termos psicológicos podem e devem ser parafraseados em termos de teorias e termos oriundos de teorias físicas. As teorias físicas que descrevem o mundo seriam a descrição completa e privilegiada da realidade, tornando as descrições biológicas e psicológicas apenas variações da descrição do mundo físico.

Essa versão específica de materialismo, baseada no projeto reducionista, é chamada de *fisicalismo*. Não se trata de negar a realidade do mental ou do psicológico afirmando que esse seria apenas uma manifestação dissimulada da atividade cerebral,

como sustentam os partidários da teoria da identidade. O reducionista *parte* da existência do mental, afirmando que esse, em última análise, é um tipo de realidade física. Ele não afirma que o mental *é* o cerebral, nem tampouco que o mental é *dispensável*, como o fazem os partidários do materialismo eliminativo – que examinaremos no capítulo V. O fisicalismo é compatível com a teoria da identidade, mas à equação "estados mentais = estados cerebrais" ele acrescentaria mais um termo: "estados mentais = estados cerebrais = estados físicos".

O reducionismo não se contenta simplesmente com a proposição de que fenômenos mentais são ou correspondem a algum tipo de mecanismo cerebral. Seu projeto fisicalista propõe uma redução mais radical, na qual o que importa são as propriedades físico-químicas dos fenômenos que ocorrem no cérebro. Em outras palavras, a natureza dos fenômenos mentais seria explicada pela natureza de algumas substâncias químicas que estariam presentes no cérebro, em especial algumas proteínas e as macromoléculas que as compõem. Essas reações e propriedades químicas podem ser específicas do cérebro, mas são, em última análise, explicáveis através de teorias físicas. Do comportamento dessas macromoléculas seria possível então retroceder, indo em direção contrária, até se poder explicar a natureza de fenômenos mentais mais complexos, como é o caso da cognição e da linguagem.

Seria muito fácil caricaturar esse projeto e muitos filósofos já o fizeram. Se considerarmos, por exemplo, que cerca de 70% de nosso tecido cerebral é composto de água, o antirreducionista poderia perfeitamente alegar que esse tipo de projeto visa mostrar que nossos pensamentos, remorsos, rancores e sonhos nada mais são do que água. O antirreducionista poderia então perguntar por que a água que está no cérebro pensa e a água que está num jarro não. Ele poderia alegar também que, se levarmos a sério esse tipo de projeto, acabaríamos explicando por que Joana D'Arc acabou queimada numa fogueira durante a Idade Média simplesmente enunciando um conjunto de leis físicas que mostram como e por que a combustão da lenha pode ocorrer.

Apesar de todas essas caricaturas filosóficas, precisamos saber como e por que esse projeto é tão levado a sério hoje em dia – talvez mais do que em qualquer outra época. As razões para isto encontram-se nos avanços da neurociência que, cada vez mais, abrem a perspectiva de explicar o funcionamento cerebral a partir de sua estrutura química.

Retomemos por um momento o que dissemos acerca de neurônios e suas conexões, as sinapses, no capítulo I. Vimos que os cérebros são constituídos de bilhões de neurônios e trilhões de sinapses. Vimos também que os neurônios têm diferentes tipos de ramificações, os dendritos e os axônios. Os axônios são muito finos e mais difíceis de serem vistos a olho nu, requerendo, frequentemente, um microscópio para poderem ser identificados. Já os dendritos correspondem a ramificações mais robustas, que podem ser vistas mais facilmente. Os dendritos, por sua vez, terminam em outras ramificações mais finas. Neurônios geram corrente elétrica; os dendritos funcionam como

uma espécie de estação receptora de sinais elétricos, como uma espécie de porto no qual são desembarcados vários tipos de mercadorias procedentes de diversos lugares[4]. As mercadorias chegam e são despachadas; quando isso acontece, o sinal que chegou no dendrito é passado para o neurônio que, dependendo da intensidade desse sinal, pode gerar ou não um novo sinal que é reenviado através do axônio. Os axônios levam então o sinal elétrico para um outro neurônio que pode estar próximo ou distante no circuito, estabelecendo o que se chama comunicação neuronal.

A comunicação neuronal, na qualidade de transmissão de carga elétrica, ocorre a partir do movimento de quatro tipos de íons: sódio, potássio, cloro ou cálcio. Os íons são mantidos dentro (potássio) ou fora do neurônio (sódio, cálcio, cloro) pela membrana, que funciona como uma barreira com duas camadas entremeadas por um tipo de gordura. Esse tipo de gordura, juntamente com outras proteínas, regula os íons que podem entrar e sair do neurônio. O interior do neurônio é sempre negativo em relação a seu exterior, gerando uma diferença de potencial ou uma voltagem (normalmente –70 ou –80 milésimos de volt).

Contudo, para que a comunicação neuronal se estabeleça, é preciso que essa barreira seja interrompida, o que é feito por algumas proteínas. As proteínas formam uma espécie de "ponte química" que permite aos íons deslocar-se para dentro e para fora do neurônio, "pulando" a camada de gordura. Essa "ponte" é chamada de canal. Para que um neurônio envie um sinal elétrico é preciso que íons positivos de sódio entrem nele, gerando uma diferença de potencial temporária: o interior do neurônio tem de se tornar mais positivo do que seu exterior (despolarização). Mas assim que a voltagem se torna positiva, íons (positivos) de potássio saem do neurônio tornando a voltagem ainda mais negativa do que o normal (hiperpolarização). Quando isto ocorre, gera-se um pulso positivo, que é seguido por um período refratário no qual o interior do neurônio mantém-se negativo por um certo tempo.

O sinal, uma vez gerado, é enviado para outro neurônio através do axônio, ou seja, os neurônios começam a se comunicar uns com os outros, repassando sinais. O contato estabelece-se através das sinapses; dendritos podem formar sinapses com outros dendritos, axônios com outros axônios etc. O tipo mais comum de sinapse é aquela na qual o axônio de um neurônio, através de seus terminais, estabelece contato com os dendritos de outro neurônio.

Suponhamos agora que temos um impulso elétrico chegando à extremidade do axônio, que se excita e torna-se temporariamente positivo. Para onde deve prosseguir esse sinal? E como? A resposta mais simples seria dizer: ele prossegue através das sinapses. Mas não é isto o que ocorre. Há uma interrupção entre um neurônio e outro e sua comunicação é feita por via química, através da ação dos *neurotransmissores*. Quando um potencial de ação, ou seja, um sinal elétrico, chega ao terminal de um axônio,

4. A metáfora e a descrição são de Greenfield (1997).

uma substância é liberada: a acetilcolina. A acetilcolina fica em pequenas bolsas no axônio e quando chega um potencial de ação, esse sinaliza que essas bolsas devem ser abertas e essa substância ser liberada. O sinal elétrico é transformado num sinal químico: quanto maior a intensidade do potencial de ação, mais bolsas de acetilcolina são abertas. A acetilcolina atravessa a sinapse rapidamente, mas, para que a comunicação entre os neurônios se estabeleça, é preciso que o sinal químico seja transformado novamente em sinal elétrico. Para que isto ocorra é preciso, por sua vez, que o neurotransmissor encontre, no neurônio seguinte do circuito, proteínas especiais chamadas de *receptores*. A presença de receptores em outros neurônios indica o caminho que os transmissores (ou neurotransmissores) devem seguir e quais os neurônios seguintes com os quais a comunicação deve ser estabelecida.

Esse processo é descrito como *lock and key* ou chave e fechadura: transmissores "procuram" receptores nesse processo de comunicação química e elétrica entre os neurônios – bilhões deles – e as sinapses que compõem nosso cérebro. A transmissão química, muito mais complexa e lenta do que seria a transmissão elétrica pura e simples a partir de circuitos fixos, dota nosso cérebro de uma enorme versatilidade: há vários tipos de transmissores que podem ser liberados em dosagens diferentes, o que permite uma grande variação na comunicação neuronal que pode ser estabelecida num determinado momento. Quando o neurotransmissor se liga ao receptor do neurônio seguinte, ativa uma série de processos no interior desse. Ativam-se mensageiros que podem ordenar que se abram canais para que a troca de cargas ocorra ou podem ir até o núcleo celular, onde estão genes que podem ser ativados para formar novos receptores na parede do neurônio. Os receptores da parede do neurônio são constantemente trocados, estando sujeitos às ordens dos genes de cada neurônio e essas ordens estão sujeitas, por sua vez, ao tipo de influência que os mensageiros exercem sobre eles.

Não sabemos ainda quais são os princípios últimos que regem essa sinfonia molecular, ou seja, por que determinados neurônios ou grupos de neurônios estabelecem comunicação entre si num determinado momento nem tampouco o que faz com que essa comunicação possa ser interrompida num momento seguinte. Ou seja, não sabemos o que determina que alguns circuitos – e não outros – façam-se e desfaçam-se o tempo todo no nosso cérebro. Espera-se que o avanço da bioquímica e da biologia molecular possa nos ajudar a esclarecer o que rege essa sinfonia cerebral: essa é a aposta do reducionista ou do fisicalista de que falamos acima.

Não há dúvida de que quando examinamos o funcionamento do cérebro e sua estrutura química mais profunda, no que se refere a mensageiros e receptores, somos levados a pelo menos concluir pela existência de uma grande dependência entre essa estrutura química e aquilo que podemos pensar. Afinal, se aquilo que chamamos de pensamento depende da comunicação entre neurônios ou é um tipo de comunicação entre eles e, se essa, por sua vez, depende de um encaixe entre mensageiros e receptores, somos levados a concluir que esses são, em última análise, os responsáveis pelos caminhos ou conexões que podem se estabelecer entre os neurônios, ou, se quisermos, os responsáveis por aquilo que *poderíamos pensar*. O que poderíamos pensar dependeria, assim, não só dos neurônios como também das reações químicas que poderiam se estabelecer entre eles.

Essa visão é extremamente tentadora e é sobre ela que se baseiam quase todos os anseios dos partidários do fisicalismo reducionista. O reducionista, contudo, enfrenta vários tipos de problemas. Em primeiro lugar, ele enfrenta o chamado *problema da tradução*. Que tipo de isomorfismo haveria entre essas reações químicas e o pensamento que permita supor a existência de algum tipo de correspondência entre essas duas entidades? Ou, em outras palavras, mesmo que para cada pensamento (cada *token*) for possível identificar uma conexão específica entre um determinado conjunto de neurônios produzida por uma reação química também específica, como poderíamos encontrar uma *passagem inteligível* entre as propriedades da reação química e as propriedades do pensamento, ou seja, aquilo que compõe seu *conteúdo* específico? O reducionista precisa de algo mais do que simplesmente estabelecer correlações, é preciso que essas se tornem *inteligíveis*. Ou seja, ele precisa saber como e por que sinais elétricos e reações químicas no nível cerebral produzem, do ponto de vista introspectivo, remorsos, lembranças e sensações que são vivenciadas subjetivamente. Encontramos, novamente, o problema da passagem da primeira para a terceira pessoa a que já nos referimos no capítulo I, desta vez sob a forma do chamado *problema da tradução*: como traduzir um manual de neurociência para um manual de psicologia? Encontrar essa "pedra da roseta" é o sonho que muitos neurocientistas contemporâneos acalentam.

Há vários aspectos envolvidos no problema da tradução. Em primeiro lugar é preciso saber se realmente a cada conexão produzida por uma reação química específica deve corresponder um único conteúdo mental. Por que o cérebro não usaria as mesmas conexões para veicular diferentes conteúdos mentais? E, mesmo admitindo que a natureza tenha produzido uma máquina tão antieconômica assim, resta saber se, uma vez encontrado um manual de tradução, esse seria suficiente para saber o que alguém está pensando. Suponhamos que algum Champollion[5] da neurociência tenha encontrado esse manual, sua pedra da roseta e, a partir dessa, tenha construído um cerebroscópio, ou seja, um aparelho que permitiria a leitura do pensamento quando acoplado ao cérebro de um indivíduo qualquer. O problema seria saber se a observação do cérebro esgota o que um indivíduo pode estar pensando num determinado momento. Será que a descrição do cérebro inclui a descrição dos fatores ambientais que podem estar compondo os conteúdos de pensamento? Ou, em outras palavras, até que ponto conteúdos mentais não envolveriam um aspecto relacional com o meio ambiente que se sobreporia à sua determinação como resultante de um conjunto de ligações químicas estritamente interno ao cérebro?

Do ponto de vista da neurociência estados depressivos estão correlacionados com a queda nos níveis de certos neurotransmissores no cérebro. Estados depressivos, entretanto, estão também correlacionados com outros fatores, como, por exemplo, mu-

5. Champollion foi o decifrador dos hieróglifos. A decifração teria se tornado possível quando ele descobriu a "pedra da roseta", um fragmento com inscrições semelhantes, mas em diferentes idiomas.

danças ambientais. Suponhamos que o fato de minha sogra ter se mudado para minha casa esteja associado com o aparecimento de minha depressão. Consulto um psiquiatra que me prescreve antidepressivos – drogas que restabelecem o nível de certos neurotransmissores no meu cérebro. Essas drogas atuam imediatamente no meu cérebro, mas, curiosamente, seus efeitos psicoterápicos levam alguns dias para começar a se manifestar. A depressão desaparece, mas apenas temporariamente. Três semanas depois ela retorna, apesar de eu continuar a tomar essas drogas regularmente. Ora, por que a droga torna-se inócua? E por que a depressão retorna apesar dos níveis de neurotransmissores terem sido restabelecidos?

A primeira resposta que surge é: posso continuar a tomar os antidepressivos, eles não mais farão efeito precisamente porque minha sogra continua a morar na minha casa. Os remédios se tornaram inócuos porque não conseguiram contrabalançar as condições ambientais, ou seja, essas se sobrepuseram na determinação de meus conteúdos mentais, apesar das ligações químicas ocorrendo no meu cérebro terem sido normalizadas. Ora, o que será então a causa de minha depressão, o desequilíbrio químico de meu cérebro ou um fator fenomênico, no caso, o fato de minha sogra ter se mudado para minha casa e continuar lá? Se for o segundo fator, isto revela que a determinação dos estados mentais pelas ligações químicas que ocorrem no cérebro não é algo tão direto e nem tão automático. Talvez essas ligações químicas contribuam apenas com parte da determinação dos conteúdos mentais e não com a sua totalidade como o reducionista gostaria. O *significado* de minha sogra ter se mudado para minha casa seria a causa de minha depressão. Outras pessoas, em circunstâncias semelhantes, poderiam não entrar num estado depressivo. Alguém poderia dizer: mas o significado de sua sogra ter se mudado para sua casa alterou o seu cérebro e produziu a depressão. Essa é, possivelmente, uma hipótese correta. O problema é que ela não explica a relação entre o significado e a alteração cerebral – ela não explica a natureza da passagem de fenômenos mentais para fenômenos cerebrais ou como esses podem se alterar mutuamente. O reducionista teria dado uma grande volta para chegar exatamente onde estava, ou seja, no ponto de partida do principal problema da filosofia da mente. Retornamos ao problema da tradução ou para o que foi chamado pelos filósofos da mente contemporâneos depois de Levine (1983) de *problema do hiato explicativo* ou *explanatory gap*.

A questão envolvida no *explanatory gap* é aquela à qual já aludimos: mesmo que alguém encontre os correlatos neurais do pensamento, isto ainda assim não explica como se passa desses correlatos neurais para características específicas que constituem um determinado conteúdo mental. Haveria mais coisas entre o céu e a terra do que nossa vã bioquímica poderia explicar. O *explanatory gap* torna-se mais evidente quando se tenta dar um passo a mais e se busca explicar, por exemplo, a natureza dos estados mentais *conscientes* em termos de seus correlatos neurofisiológicos. Ou seja, busca-se uma conexão forte, inteligível e explicativa entre eventos neuronais e eventos conscientes. Mas nenhum elemento físico presente no cérebro *implica* na produção da consciência. Poderíamos ter um sistema que reproduz todas as características físicas do cérebro e, ainda assim, esse sistema poderia não produzir estados conscientes.

A ideia que está por trás do *explanatory gap* é a de que não há nenhuma característica física específica que possamos atribuir a estados subjetivos tais como a percepção de dores, cores etc., e tampouco nenhuma característica física de um estado mental qualquer que o torne um estado consciente. Contudo, temos acesso subjetivo a esses estados como sendo estados conscientes, por isso sabemos que eles existem. Como eles não têm uma característica física, não seria possível montar uma história causal que nos leve de estados cerebrais a estados conscientes – essa história sempre pressuporia, em algum momento, algum tipo de salto ou passagem não explicada. O mesmo se aplicaria ao caso das nossas sensações: a descrição dos fenômenos bioquímicos e cerebrais subjacentes à dor seriam insuficientes para determinar o que é *sentir dor*. Esse aspecto subjetivo não seria capturado por essa descrição, tornando-a insuficiente para sabermos como o cérebro gera estados subjetivos. O que estaríamos encontrando aqui – esse salto não explicado – corresponde, mais uma vez, à passagem da perspectiva de terceira pessoa para uma perspectiva de primeira pessoa do qual falamos no capítulo I.

Esse salto corresponderia também a alguma quintessência que, segundo Descartes, ficaria faltando se pudéssemos construir uma réplica perfeita de um ser humano na forma de um autômato, pois, na medida em que ser consciente não seria uma propriedade física, a replicação física integral de um cérebro não implicaria, necessariamente, na replicação do caráter consciente dos estados mentais que esse autômato poderia vir a ter. Esses estados subjetivos conscientes seriam aquilo que os filósofos da mente batizaram com o nome de *qualia* e seriam, em última análise, inescrutáveis do ponto de vista de uma descrição física. Alguns filósofos da mente como Dennett rejeitam a própria existência dos *qualia* e sustentam que eles nada mais são do que uma excrescência metafísica ou linguística que só faria sentido se aceitássemos todos os pressupostos da metafísica cartesiana, sobretudo o dualismo entre mente e cérebro.

Antes de enveredarmos por essa discussão – que deixaremos para o capítulo seguinte – precisamos tentar responder uma questão mais urgente para o reducionista: será esse salto explicativo resultado de uma intransponibilidade real que encontraríamos nas nossas tentativas de passar de cérebros para mentes e dessas para estados conscientes ou será ele apenas uma condição provisória dos nossos conhecimentos acerca das relações entre mente e cérebro, algo que podemos esperar ser alterado no futuro? Em outras palavras, será o *explanatory gap* inevitável por ser uma característica do mundo ou algo que resulta apenas do modo como estruturamos nosso conhecimento e nossas explicações científicas? Discutiremos rapidamente apenas essa segunda possibilidade, mas veremos, a partir dessa discussão que, mesmo que o *explanatory gap* resultasse apenas do modo como estruturamos nossas explicações científicas, ainda assim ele seria um sério obstáculo para o reducionista.

A ideia de redução como mecanismo para explicação, sobre a qual se apoia o projeto do fisicalismo na sua tentativa de trilhar um caminho da biologia molecular até a consciência, parece apoiar-se num princípio contraintuitivo: o de que poderíamos explicar o funcionamento de um motor a combustão pela estrutura atômica das partícu-

las que compõem esse motor[6]. O caráter contraintuitivo desse exemplo deriva-se do fato de que, embora os componentes do motor estejam sujeitos às leis subatômicas que regem as partículas elementares, esse tipo de conhecimento seria irrelevante para explicarmos o funcionamento desse mecanismo. Em outras palavras, quando um motor funciona, embora estejam ocorrendo eventos em nível subatômico o tempo todo, deles não podemos derivar uma explicação para o funcionamento desse mecanismo: haveria um hiato entre esses dois tipos de níveis de explicação.

Putnam (1973) sugere que aqui encontramos não apenas um hiato, mas uma incompatibilidade. Ele nos convida a imaginarmos um tabuleiro feito de plástico, que teria dois furos: um quadrado e um redondo. Temos também um pino grande, de metal, arredondado. O pino passa pelo furo redondo, mas não passa pelo quadrado, que é menor: podemos simplesmente observar isto. A explicação do motivo por que o pino não passa pelo furo quadrado é simples e intuitiva: no máximo, precisaríamos invocar algumas leis do comportamento físico macroscópico dos objetos que estamos considerando.

Mas o que ocorreria se optássemos por elaborar uma explicação desse tipo de fenômeno em nível microscópico, ou seja, no nível das partículas subatômicas que compõem esses objetos? Esses objetos passariam a ser representados como nuvens de elétrons e prótons ou como os átomos que os compõem, formando três nuvens. Ora, ocorre que se os concebemos assim, em nível microscópico, o pino grande pode passar pelo buraco quadrado, o que não ocorre em nível macroscópico quando os consideramos como objetos rígidos e visíveis, pura e simplesmente. A explicação em nível microscópico não pode explicar aquilo que vemos habitualmente e, embora esses três objetos possam receber uma descrição subatômica, esse tipo de descrição é absolutamente irrelevante quando queremos saber por que o pino não passa pelo buraco quadrado do tabuleiro. Qualquer objeto pode ser descrito macroscópica e microscopicamente, mas, em alguns casos, se montarmos nossas explicações a partir da descrição microscópica podemos obter relatos contraintuitivos ou irrelevantes.

O mesmo poderia ser dito se tentássemos explicar fenômenos sociais e econômicos apelando para uma descrição microscópica da biologia e da neurofisiologia dos agentes humanos envolvidos nesse tipo de fenômenos. Chegaríamos rapidamente à tentativa de explicar o assassinato de Joana D'Arc através das leis da combustão da lenha a que nos referimos no início deste capítulo. Em outras palavras, quando montamos explicações, é preciso preservar uma certa independência dos níveis de explicação em que nos situamos, caso contrário forjaremos teorias ridículas. A existência de um hiato entre esses níveis de explicação – um *explanatory gap* – parece ser inevitável.

Ora, será então o *explanatory gap* um problema incontornável para o projeto do reducionismo fisicalista? Uma maneira de tentar contornar os problemas colocados pelo *explanatory gap* seria apostar na produção das chamadas propriedades emergen-

6. Objeção semelhante encontramos em Searle (1997).

tes. Eu poderia, por exemplo, supor que a consciência é uma propriedade emergente em relação a um determinado fenômeno físico ou um determinado tipo de organização funcional observável no cérebro ou na natureza. Essa propriedade não seria redutível a cada um dos componentes dessa organização, mas eu poderia explicá-la sem ter de, necessariamente, romper com uma proposta fisicalista. Por exemplo, a solidez do gelo é uma propriedade física desse tipo, uma propriedade que não pode ser reduzida a cada uma das moléculas de água que compõem uma barra de gelo. Entretanto, para explicar a solidez do gelo não preciso romper com leis físicas ou apelar para propriedades ocultas ou inescrutáveis. Analisaremos essa proposta a seguir.

Emergentismo e superveniência

As chamadas teorias da *superveniência do mental sobre o físico* acenam com a possibilidade de se sustentar o materialismo, evitando, contudo, os problemas oriundos de sua versão reducionista. A teoria da superveniência sustenta que existe uma relação de dependência entre todos os fenômenos do universo e sua base física. Fenômenos biológicos, por exemplo, dependem de sua base física na medida em que os seres vivos são compostos de elementos físicos – moléculas de carbono, nitrogênio etc. – e uma alteração biológica seria inconcebível sem ser acompanhada por uma alteração correspondente no seu nível físico. Em outras palavras, a teoria da superveniência estipula uma dependência entre os diversos níveis que podemos identificar na observação de um determinado fenômeno. A relação de dependência não se aplica unicamente a fenômenos, mas também às *propriedades* dos fenômenos.

De um modo genérico a noção de superveniência consiste na formalização da ideia de que um conjunto de fatos e propriedades pode determinar integralmente um outro conjunto de fatos ou propriedades: fatos e propriedades B (de nível mais alto) supervêm sobre fatos e propriedades A (de nível mais básico) quando:

a) Existe uma covariação entre fatos e propriedades B e A.

b) Existe uma dependência entre fatos e propriedades B e A.

c) Fatos e propriedades B não são redutíveis a A.

A relação de superveniência é uma das noções mais complexas introduzidas pela filosofia da mente contemporânea e não poderemos discuti-la aqui em todos os seus detalhes. Antes de aplicá-la especificamente à relação mente e cérebro, nós a ilustraremos através de exemplos. Todos eles ilustram a aplicação das cláusulas a), b) e c), mas têm diferenças importantes entre si. Os dois primeiros apelam para uma noção mais intuitiva e cotidiana de superveniência onde a irredutibilidade de uma propriedade superveniente em relação à sua base física é identificada com a ideia de uma *propriedade emergente*. O terceiro exemplo mostra, entretanto, que nem todas as propriedades supervenientes configuram-se, necessariamente, como propriedades emergentes.

Primeiro exemplo: Suponhamos que você esteja olhando para um quadro, digamos, a Mona Lisa. Certamente você vai querer atribuir a esse quadro a propriedade de

"ser belo". Ora, esse quadro não poderia ser belo se ele não fosse constituído de elementos físicos, como, por exemplo, ser um pedaço de tela sobre a qual foram aplicadas pinceladas coloridas. Se destruirmos a tela, destruiremos o quadro e estaremos destruindo também a possibilidade de que esse quadro seja visto como algo belo. Por outro lado, se alterarmos as cores e as pinceladas que estão impressas na tela, estaremos alterando o quadro e, muito possivelmente, ele deixará de ser belo. Com isto, estamos estipulando que "ser belo" depende de uma base física e, ao mesmo tempo, que se a base física variar, a propriedade "ser belo" variará também, ou seja, estipulamos uma covariação entre ser belo e a alteração das propriedades físicas do quadro, no caso de estarmos mudando ou acrescentando novas pinceladas à tela original. "Ser belo", não é, entretanto, uma propriedade que seja redutível ou igualável a um pedaço de tela, nem tampouco a cada uma das pinceladas que estão impressas sobre ela. Não são as pinceladas que são belas e sim o quadro de Leonardo da Vinci. "Ser belo" é uma propriedade que supervém à tela e às pinceladas – e sem essa propriedade superveniente não poderíamos falar nem de Mona Lisa nem tampouco que a Mona Lisa é bela.

Segundo exemplo: Sabemos que a água, se refrigerada a uma temperatura inferior a 0 graus centígrados, torna-se gelo. Passa-se do estado líquido para o estado sólido. As propriedades da água no estado sólido são diferentes da água em estado líquido. A solidez e a impenetrabilidade são exemplos de propriedades que ocorrem quando a água encontra-se em estado sólido – propriedades que não são comuns ao estado líquido. Estaremos aqui diante de uma relação de superveniência? É provável que sim. "Ser sólido" não existiria sem uma base física; é a variação dessa base física (refrigerar a água) que possibilita o aparecimento dessa propriedade, ou seja, há uma covariação entre "tornar-se sólido" e alterar a temperatura da água. Será "ser sólido" o resultado da alteração de cada um dos átomos da água? É bem provável que, para produzir a solidez, cada um dos átomos da água tenha de sofrer uma alteração. Entretanto, "ser sólido" não parece ser uma propriedade que poderia ser aplicada individualmente a cada um dos átomos da água, pois não parece fazer sentido dizer que "um átomo é sólido", embora cada um deles concorra para a produção da propriedade "ser sólido". Nesse sentido específico, solidez é uma propriedade superveniente da água quando essa é transformada em gelo.

Terceiro exemplo: Suponhamos que você esteja observando as ações de uma pessoa. Do ponto de vista ético, podemos considerar as ações dessa pessoa como "boas" ou "más". Essas ações certamente têm uma base física: correspondem aos movimentos musculares de um organismo. Se não houvesse esses movimentos musculares não haveria nada que pudéssemos estar qualificando de "bom" ou de "mau". A qualificação "bom" ou "mau" varia de acordo com o tipo de ação que for executado. Entretanto, não há nenhuma característica física dessas ações que nos permita qualificá-las de "boa" ou de "má". O aspecto moral das ações existe e, contudo, ele é irredutível a uma base física.

Os dois primeiros exemplos introduzem a ideia de superveniência ligada ao aparecimento de uma propriedade emergente e ilustram a dependência e a irredutibilidade dessa última a uma base física. O terceiro exemplo não difere muito do primeiro: em

vez de um julgamento estético, temos um julgamento ético que leva, no primeiro exemplo, ao aparecimento da propriedade "ser belo" e, no terceiro, ao aparecimento das propriedades "ser bom" ou "ser mau". Mas o primeiro exemplo e o terceiro têm uma diferença fundamental em relação ao segundo: "ser belo" ou "ser bom" não são propriedades que emergem da mesma maneira que a solidez emerge do resfriamento da água. A diferença está no fato de que a propriedade emergente "ser sólido" deriva-se *necessariamente* da alteração da base física do fenômeno em questão, qual seja, do resfriamento da água. Sempre que a água é resfriada abaixo de zero graus ela se torna gelo e a solidez aparece. O mesmo não ocorre no primeiro exemplo, onde a propriedade "ser belo" não emerge necessariamente, dependendo sempre do ponto de vista de um observador ou apreciador. Alguns observadores podem considerar a Mona Lisa bela, outros não. O primeiro e o terceiro exemplo são similares, embora possamos ressaltar, no caso do julgamento moral, uma nuança específica: o julgamento moral de uma ação está sempre condicionado pelo contexto onde essa ocorre, seja esse um contexto ambiental específico ou, mais amplamente, um contexto histórico-cultural ou valorativo que nos inclinará a considerar, em última análise, uma ação como tendo sido boa ou má. Em outras palavras, no julgamento moral não podemos considerar uma ação isoladamente. O mesmo ocorre no caso do julgamento estético, embora em menor grau: a despeito da mudança de contexto histórico e de época, sempre há alguém que continua considerando a Mona Lisa um belo quadro.

O que ocorre quando se tenta conceber a relação entre mente e cérebro como uma relação de superveniência? Uma das vantagens é poder estipular a existência de uma dependência e de uma covariação entre fenômenos mentais e fenômenos cerebrais, sem, entretanto, ter de estipular ou mostrar como uns poderiam ser *reduzidos* aos outros. Certamente a ideia de propriedade emergente concebida como propriedade superveniente pode auxiliar-nos a resolver alguns problemas colocados pelo materialismo reducionista ou pela teoria da identidade. Por exemplo, os paradoxos da localização estrita dos estados mentais no espaço desaparecem. O mental, como propriedade emergente, não precisa ser localizado nesse ou naquele ponto do espaço. Uma ideia de distribuição do mental pelo sistema nervoso como um todo pode surgir – e essa é sem dúvida uma concepção mais adequada às ideias neurológicas contemporâneas que descartam o localizacionismo estrito e inclinam-se em direção a uma concepção integracionista do funcionamento cerebral. De modo geral, livramo-nos de quase todos os paradoxos do reducionismo ao adotar a teoria da superveniência do mental em relação ao físico.

Contudo, poderíamos perguntar: em qual dos três exemplos acima se enquadra a visão do mental como superveniente ao físico? Podemos manter a ideia de propriedade emergente, como ocorre no primeiro e no segundo exemplos. Mas será a relação de superveniência entre o mental e o físico uma relação dependente de um observador ou apreciador – como no caso do primeiro exemplo? Ou será uma relação que implica uma necessidade entre base física e propriedade superveniente – como no segundo exemplo?

Se o modelo a ser adotado for o do primeiro exemplo, ou seja, que o mental como propriedade superveniente depende da identificação de um observador, somos imediatamente levados em direção à ideia de que propriedades emergentes/supervenientes estariam apenas nos olhos desse observador. Essa é uma posição cômoda para adotarmos, pelo menos à primeira vista. Contudo, nesse caso, abre-se a possibilidade de alguns observadores poderem identificar propriedades emergentes/supervenientes e outros não. Se a identificação de propriedades mentais consideradas como emergentes/supervenientes pode estar apenas nos olhos do observador e se sua identificação pode variar de um observador para outro, como e por que ocorre então essa variação? A atribuição de propriedades mentais a um fenômeno torna-se então totalmente relativizada. Podemos imaginar então uma situação tal que, para um leigo, o cérebro estaria produzindo propriedades emergentes sob a forma daquilo que ele chamaria de "vida mental". Para um neurocientista, o leque daquilo que ele consideraria propriedades emergentes seria muito menor. O conhecimento do neurocientista poderia aumentar muito, ou seja, proporcionalmente ao avanço que a própria neurociência pode nos proporcionar. No limite, não haveria mais nenhuma propriedade emergente para o neurocientista: esse teria chegado seja a alguma versão da teoria da identidade, seja a alguma versão forte de materialismo reducionista/fisicalista. Falar de emergência ou de superveniência entre o físico e o mental tornar-se-ia absolutamente desnecessário para esse neurocientista do século XXII. Mas, ao endossar alguma versão forte de teoria da identidade ou de reducionismo fisicalista não passaria ele, implicitamente, a incorrer em todas as dificuldades que apontamos para essas variedades de materialismo?

Vejamos o que ocorreria se concebêssemos nossa relação de emergência/superveniência adotando como modelo o segundo exemplo. Adotar o modelo do segundo exemplo significa ter de provar que o mental *necessariamente* supervém sobre o físico. Significa ter de provar que uma réplica física perfeita de um ser humano qualquer – seja ela produzida em laboratório ou por acaso, digamos, por um acidente quântico – replicará, *necessariamente*, a vida mental desse ser humano. É difícil poder provar que isto necessariamente aconteceria. Mostrar que existe relação de necessidade entre a produção de uma réplica física de um ser humano e a produção de uma vida mental correspondente equivaleria a sepultar definitivamente o argumento cartesiano em favor da existência de uma "quintessência" que nos distinguiria dos autômatos – o argumento que apresentamos no capítulo anterior. Significaria também resolver a questão do *explanatory gap* de que falamos na seção anterior. Alguém que fizesse isto mereceria mais do que um Prêmio Nobel.

Como ganhar prêmios Nobel é muito difícil, mas não impossível, talvez tenhamos de abordar essa questão a partir de outro ângulo: haveria argumentos *contra* a possibilidade de se mostrar que a vida mental necessariamente emergeria de uma réplica perfeita e integral de um ser humano ou de um cérebro humano? Essa maneira de formular o problema nos encaminha, de certa maneira, para as dificuldades que poderiam surgir da ideia de conceber a relação de superveniência a partir do terceiro exemplo. Ora, aqui esbarramos num problema sério logo de início: ser "bom" ou "ser mau",

como propriedade moral e superveniente de algum tipo de ação, pode depender do contexto no qual essa ação for executada. Todos nós sabemos disto: uma boa ação pode ser má num contexto e circunstância diferentes. Nesse caso, a replicação de propriedades físicas torna-se claramente insuficiente. Outros exemplos podem tornar essa afirmação ainda mais clara[7]. Se replicarmos uma escultura de Rodin, não estaremos replicando todas as suas propriedades, pois teremos produzido uma cópia, que certamente valerá muito menos do que a original. A situação torna-se pior ainda se replicarmos notas de 100 reais. Essas não só não terão o valor de uma nota de 100 reais, como, pior ainda, serão uma falsificação. Gera-se uma propriedade indesejada e nenhuma daquelas que poderíamos atribuir a uma nota de 100 reais autêntica.

Pensemos agora em algumas teorias da mente que sustentam que a vida mental formou-se, ao longo do processo evolucionário, como estratégia adaptativa de um tipo especial de mamíferos, a saber, os seres humanos. Ora, dois organismos podem ter exatamente as mesmas características físicas, mas, se um deles se encontrar num ambiente muito diferente do outro, a replicação das características físicas não será condição suficiente para dizermos que ambos serão igualmente adaptados. Um deles poderá ser adaptado e outro não. Ora, se tomamos como hipótese que a consciência se formou como estratégia de adaptação (adotando, por exemplo, uma posição neodarwiniana acerca do aparecimento da consciência e da experiência consciente), isto pode nos levar à ideia de que dois seres com exatamente as mesmas características físicas poderiam diferir no que diz respeito a ter experiências conscientes ou não. Obviamente, nesse caso, a *história evolucionária* deles seria diferente, e poder-se-ia argumentar que é esse fator que faz toda a diferença. Contudo, isto poderia nos levar, igualmente, à bizarra conclusão de que dois organismos, fisicamente indiscerníveis, teriam de ser classificados como membros de espécies diferentes.

A esta altura só resta ao partidário da teoria da superveniência expandir essa noção. Ele terá de falar em superveniência *local*, como designando a superveniência/emergência produzida por uma replicação física, e em superveniência *global*, designando a replicação física acompanhada de uma replicação conjunta de circunstâncias e contextos ambientais. A nova formulação de seus princípios, agora ampliada, terá a seguinte forma: existe uma covariação e uma dependência entre fatos e propriedades B e A. Isto é, fatos e propriedades B, de nível mais alto, supervêm sobre fatos A de nível mais baixo (físico). A replicação das características físicas de A é necessária, mas não suficiente, para a produção de B se na produção de B estiverem envolvidos fatores contextuais e circunstanciais. Para salvar a sua teoria, o partidário da superveniência introduz então uma cláusula *ad hoc*: replicando fisicamente A, replica-se todas as propriedades de B na qualidade de propriedades supervenientes *se as características globais do mundo onde ocorre A forem mantidas*. O princípio de superveniência passa então a ser generalizado da seguinte forma: em todos os

7. Estes exemplos são derivados de Chalmers (1996).

mundos possíveis, semelhantes ao nosso (isto é, que tenham a mesma estrutura lógica e onde as leis físicas de nosso mundo sejam igualmente aplicáveis), ao se replicar fisicamente A replica-se B. Se B for uma propriedade mental ela será superveniente a A em qualquer um desses mundos possíveis.

Ao introduzir o conceito de mundo possível com uma estrutura semelhante à do nosso, onde supostamente o mental supervém ao físico, o partidário da teoria da superveniência estaria contornando a dificuldade do contexto e da circunstância. A cláusula condicional e a possibilidade de conceber mundos possíveis estaria então salvando sua teoria ao fixar a necessidade de manter a covariação de contexto e circunstância para que a dependência do mental em relação ao físico se mantenha. Poderíamos afirmar, então, que a ideia de mundo possível entra nessa história na qualidade de uma hipótese auxiliar, como aquelas de que lançamos mão quando queremos fazer uma demonstração matemática ou geométrica. A solidez do gelo será emergente e superveniente em todas as circunstâncias ou mundos possíveis onde a água seja H20 e onde ela se solidifique quando resfriada a 0 graus centígrados.

A introdução desse conceito de superveniência global, exigindo a covariação de circunstâncias, contextos, leis lógicas e leis físicas que devem ser hipotetizadas e exportadas para um mundo possível para podermos afirmar que nesse mundo o mental continuaria sendo superveniente ao físico, pode nos levar a um campo minado. Saberíamos por que somente a Monalisa pintada por Leonardo da Vinci realmente vale muito dinheiro, o mesmo não se aplicando a suas réplicas produzidas posteriormente. Poderíamos também explicar por que a nota de 100 reais seria uma falsificação, dizendo que só aquelas produzidas pelo Banco Central têm valor e que foi esse elemento contextual que faltou na replicação. Por outro lado, poderíamos questionar se uma réplica física da Madame Tatcher preservaria sua propriedade de "ser uma política". Supostamente, pela teoria da superveniência, uma réplica física da Madame Tatcher replicaria também sua vida mental, seus pensamentos e seus discursos. Seria isto suficiente para preservar a propriedade "ser uma política"? O que ocorreria se essa réplica física da Madame Tatcher fosse concebida num mundo possível igual ao nosso, onde leis físicas e lógicas se mantenham, como por exemplo, a Inglaterra há 200 anos atrás? Certamente a propriedade "ser política" não seria mantida como propriedade superveniente e essa réplica não seria sequer vista como sendo Madame Tatcher, mas como uma louca fazendo discursos estranhos[8]. Ora, a propriedade "ser política" só poderia ser mantida se reproduzíssemos, nesse mundo possível, um mundo exatamente igual àquele em que Madame Tatcher viveu e foi primeira-ministra da Inglaterra. Assim sendo, a noção de "mundo possível" torna-se, no limite, dispensável ou inócua. O que sobra da tese da superveniência global é que o mental e todas suas propriedades supervêm à réplica de sua base física se essa réplica estiver situada exatamente no mesmo contexto, circunstância, horário etc., do mundo em que vivemos. Como se esses con-

8. Este exemplo é fornecido por Smith (1993).

textos, circunstâncias etc., fossem passíveis de uma descrição completa. Teríamos de supor um mundo possível que tivesse todas as mesmas características físicas do nosso, incluindo também toda a sua história desde sua formação até os anos de 1980 para então, lá, podermos situar a réplica da Madame Tatcher e poder dizer que a propriedade "ser política" supervém à sua base física. O partidário da teoria da superveniência teria de herdar o universo. Ou então arranjar uma maneira melhor de definir "mundo possível". Um bom começo seria saber, acerca desses mundos possíveis, a que horas eles abrem, a que horas eles fecham e a que distância eles ficam do centro da cidade.

O futuro do materialismo: a noção de matéria

O filósofo inglês Bertrand Russel (1872-1971) no seu livro *The analysis of matter* [A análise da matéria] (1927) sugere que nunca podemos observar diretamente nosso cérebro. Não temos acesso direto ao cérebro, mas apenas a *representações* desse. Ele nos diz que "quando um neurofisiólogo olha para um cérebro, o que ele vê é uma parte de seu próprio cérebro e não parte do cérebro que ele está examinando"[9]. Em outras palavras, Russell nos diz que nosso cérebro cria imagens de outros cérebros quando os examina; só temos acesso a essas imagens ou a essas representações. Nosso conhecimento do cérebro – seja o nosso ou de outrem – é mediado pela nossa própria percepção. Cérebros são parte do mundo material e não temos acesso direto à matéria, mas apenas a percepções ou representações desta.

Podemos não concordar com Russell, mas sua observação aponta para uma grande dificuldade do materialismo: ter tratado o conceito de matéria de forma ingênua. Em outras palavras, o grande problema do materialismo não seria reduzir o mental ao material, mas, antes de mais nada, em saber o que seria a matéria. Esse lado do problema não pode ser esquecido e talvez seja o aspecto mais espinhoso a ser enfrentado pelo materialista, ao contrário do que poderíamos supor à primeira vista. Não sabemos o que é a matéria, e mesmo que o materialista sustente que não está reduzindo o mental ao material e sim a uma descrição do mundo material feita pela ciência – como é o caso do fisicalismo –, ainda assim, ele enfrenta a falta de um consenso acerca do que a ciência física pode nos dizer acerca desse assunto.

Em 1951 o físico Erwin Schrödinger, num pequeno ensaio intitulado *Science and Humanism* [Ciência e humanismo], já sugeria que quando fazemos a pergunta "O que é a matéria?" estamos nos perguntando, na verdade, "Qual é nosso esquema *mental* da *matéria*?" Até o início do século XX a matéria parecia ser algo permanente, algo sólido que poderíamos agarrar com nossas mãos. Segundo essa concepção, a matéria estava inteiramente submetida a um conjunto de leis a partir das quais, dadas as condições iniciais, seria possível predizer seu comportamento.

Contudo, essa concepção tradicional foi radicalmente modificada: não concebemos mais a matéria como algo tangível no espaço ou algo que segue leis totalmente predeterminadas. A matéria é composta de átomos, mas já não se pode concebê-los como se esses fossem pequenos corpúsculos identificáveis e tangíveis como os corpos que nos rodeiam. Ao iniciar-se o século XX, foram descobertos vários fenômenos acerca da estrutura dos átomos que contrariaram essa concepção tradicional. Em vez de serem partículas sólidas e duras descobriu-se que os átomos consistiam de imensas regiões de espaço nas quais partículas moviam-se em torno de um núcleo: um objeto extremamente pequeno, mas enorme quando consideramos a distância entre seu núcleo e as partículas que o rodeiam. O aparecimento da teoria quântica mostrou, ademais, que as entidades subatômicas da matéria são extremamente abstratas e dotadas de um aspecto dual; manifestando-se às vezes como partículas e às vezes como ondas – uma natureza dual também exibida pela luz, que pode assumir a forma de onda eletromagnética ou de partícula.

Schrödinger, no seu ensaio, ressalta ainda um aspecto desconcertante dessa nova concepção de matéria: seus componentes não possuem nenhum tipo de "identidade". Quando se observa uma partícula qualquer, por exemplo, um elétron, devemos considerar isto como um *evento singular*, ou seja, algo que não se repetirá mais. Se observamos, pouco tempo depois, uma partícula similar, num ponto muito próximo ao primeiro, e até se estipulamos a existência de uma relação causal entre a primeira e a segunda observação, não podemos, contudo, afirmar que se observa a mesma partícula no primeiro e no segundo caso: nada garante a existência de uma "identidade" entre as partículas observadas em ocasiões diferentes.

O que ocorre quando observamos um corpo visível e palpável diante de nós? Um átomo é composto por várias partículas; vários átomos formam uma molécula e todos os objetos visíveis e palpáveis que nos rodeiam são compostos de moléculas. Mas como conceber a individualidade e a identidade desses objetos visíveis e palpáveis se as partículas elementares que os compõem não têm individualidade nem identidade?

Para tentar responder essa pergunta, Schrödinger nos convida a considerar um caso particular: a identidade e a individualidade de uma estatueta sobre uma escrivaninha, usada como peso para segurar folhas de papel[10]. Essa estatueta foi herdada de seu pai há 50 ou 60 anos, tendo uma longa história e tendo estado em vários lugares diferentes até chegar a ficar sobre a escrivaninha. Como posso saber que essa estatueta é a mesma que foi herdada há 50 anos atrás, se suas partículas elementares não têm individualidade nem identidade?

A resposta dada por Schrödinger é que a individualidade e a identidade não são determinadas pela estrutura da matéria que compõe a estatueta, mas por uma "*Gestalt*" que tenho dela. Essa *Gestalt* ou essa *forma* que tenho da estatueta na minha mente de-

10. Este exemplo foi reproduzido de Schrödinger (1951).

termina que esta seja a mesma estatueta de 50 anos atrás. Se a estatueta tivesse sido derretida e, a partir do metal fundido, tivesse sido modelado um outro objeto qualquer, eu certamente não apostaria na sua identidade ao longo do tempo. Em outras palavras, o material de que é feita a estatueta é secundário na determinação de sua individualidade e de sua identidade.

Suponhamos que um dia, depois de velho, eu retorne para o lugar onde nasci. Olho em volta e constato que o lugar continua exatamente o mesmo: um campo, com um rio e uma pequena casa de madeira. Digo que o lugar continua o mesmo porque sua forma não mudou – a forma que tenho em minha mente. Mas certamente a água que corre pelo rio é outra, as partículas elementares que compõem a casa, os objetos que estão no campo não são mais os mesmos. Nem sequer meu próprio corpo é o mesmo. Contudo, digo que nada mudou, pois retive na minha mente a forma primitiva de organização dos objetos desse lugar.

Ao estipular que é a forma ou uma organização o que constitui a identidade dos objetos e que essa forma está na minha mente ou na minha percepção, abandono a ideia tradicional de que é o material de que são constituídos os objetos que determina sua identidade ao longo do tempo. Essa forma não precisa ser forma de algo; a forma pode existir independentemente de referir-se ou não a algum tipo de substrato material. Ora, se é a *Gestalt* ou a forma de um objeto que está na minha percepção ou na minha mente que determina a individualidade e a identidade desse objeto, e não a matéria de que ele é composto, até que ponto minha concepção de matéria não depende de minha mente?

Schrödinger, o cientista que formulou a equação fundamental da mecânica quântica, expressou esse tipo de preocupação várias vezes, inclusive num outro ensaio seu *Mind and Matter* [Mente e matéria], publicado em 1958. Werner Heisenberg, um dos fundadores da física contemporânea, expressou preocupações similares, enfatizando que as leis da natureza não mais se referem às partículas elementares, mas ao *conhecimento* que temos delas. Alguns anos mais tarde (1962), o grande físico ganhador de Prêmio Nobel, Eugene Wigner, publicou um ensaio intitulado "Remarks on the Mind-Body Question" [Observações sobre o problema mente-corpo]. Nesse ensaio, Wigner afirmou que não seria possível formular as leis da mecânica quântica de modo consistente sem referência à nossa consciência. O estudo do mundo nos devolve ao estudo de nossa própria consciência. Todo nosso conhecimento estaria preso a um tipo de círculo vicioso do qual dificilmente poderíamos escapar.

Esse círculo vicioso inicia-se com o estudo da mente humana, que passa a ser explicada a partir da atividade do sistema nervoso central. O sistema nervoso central, por sua vez, pode ser reduzido a algum tipo de estrutura biológica – uma estrutura que pode ser explicada pela física de partículas, isto é, pela interação dos átomos de carbono, nitrogênio, oxigênio e assim por diante. Finalmente, a física atômica – a mecânica

quântica – remete-nos de volta à mente ou à consciência, que deve ser tomada como constituindo um componente primitivo do sistema[11].

Se Wigner estiver certo, um eventual triunfo do fisicalismo seria uma façanha peculiar: em vez de se reduzir a psicologia à física, teríamos conseguido o indesejável resultado (ao menos para alguns físicos) de reduzir a física à psicologia. Um resultado que só pode ser evitado se alguém mostrar que essa interpretação da mecânica quântica deve ser abandonada por ser necessariamente incorreta. Se esse não for o caso e, se os conceitos de "mente" e de "consciência" tiverem de integrar o vocabulário da física, o sonho de uma ciência da natureza escrita em linguagem extensional se torna, da mesma maneira que na psicologia, um projeto inexequível.

O QUE LER

Sobre teorias da identidade:

Borst, C.V. *The Mind-Brain Identity Theory*

Rosenthal, D.M. *Materialism and the Mind-Body Problem*

Rosenthal, D.M. *The Nature of Mind*

Armstrong, D. *A Materialist Theory of the Mind*

Sobre teorias da superveniência:

Kim, J. *Philosophy of Mind*

Kim, J. *Supervenience and Mind*

Sobre o *explanatory gap*:

Verbete "consciousness" de Guttenplan, S. *A Companion to the Philosophy of Mind*

Sobre os principais tópicos abordados pela neurociência:

Greenfield, S. *The Human Brain: a guided tour*

11. Cf. a este respeito Morowitz (1980).

CAPÍTULO IV

AS VARIEDADES DO DUALISMO

A principal proposição sustentada pelo dualismo é a existência de uma diferença fundamental entre o físico e o mental, entre mente e matéria. Essa proposição choca-se frontalmente com as expectativas da ciência contemporânea, que tem buscado, incessantemente, uma explicação para a natureza dos fenômenos mentais através do estudo do funcionamento cerebral dos seres vivos. Para a ciência, nossos corpos e, sobretudo, nosso sistema nervoso são parte do mundo físico e, na medida em que alterações da mente não podem ocorrer sem uma alteração correspondente no cérebro, não haveria nenhuma razão – a não ser nossa ignorância científica provisória – para supor que mente e cérebro devam ser radicalmente distintos. Dessa perspectiva, o dualismo só seria sustentável como parte de alguma crença religiosa.

Contudo, não é nenhuma motivação religiosa explícita que tem levado alguns pensadores a defender o dualismo no século XX. Não podemos sequer falar de dualismo no século XX como sendo um movimento monolítico. Há vários tipos ou variedades de dualismo, que agruparemos em duas grandes vertentes: o dualismo de substâncias ou *dualismo substancial* e o dualismo de propriedades ou *dualismo de atributos*.

O primeiro tipo postula a existência de uma substância mental, cujas propriedades seriam totalmente distintas e incompatíveis com o mundo material. Além disso, o dualismo substancial postula que a identidade de uma pessoa, bem como sua sobrevivência após a morte, deve-se a essa substância imaterial, frequentemente chamada de "alma" ou "espírito". Esse ponto de vista, herdeiro tardio da filosofia cartesiana, encontra-se hoje em dia praticamente abandonado, por ter sido incapaz de contornar o problema de saber como uma substância imaterial poderia interagir com o corpo. Essa dificuldade inerente ao dualismo de substâncias – já notada pelos filósofos posteriores a Descartes – torna-o uma posição insustentável, na medida em

que não poderíamos explicar a correlação entre estados mentais e comportamentos em termos de um enlace causal.

O dualista de substâncias encontra-se, assim, confinado a um dilema. Ou bem ele abre mão do princípio de causalidade – o que implica rejeitar a imagem científica do mundo – ou bem ele propõe uma dissociação completa entre mente e comportamento. Ambas alternativas levam rapidamente a consequências paradoxais ou indesejáveis.

A primeira alternativa exclui a mente do domínio da ciência: nada poderíamos afirmar acerca da natureza do mental além do fato de ele ser distinto da matéria, o que torna o dualismo de substâncias uma filosofia sem agenda – conforme já assinalamos no capítulo I. A segunda alternativa – dissociar completamente mente e comportamento – parece ser ainda mais bizarra. Se mente e corpo não interagem causalmente na produção e variação do comportamento, isto implica sustentar que nossas mentes são inúteis ou desnecessárias. Nesse caso, o dualista de substâncias evitaria contradizer o princípio de causalidade, mas acabaria por contradizer outra pressuposição básica da ciência: a de que nossas mentes devem ser algo útil para a sobrevivência de nossa espécie; algo útil que teria se formado ao longo de um processo evolucionário. Obviamente, nem tudo que a evolução produz é necessariamente útil ou tem uma finalidade discernível. O melão não vem dividido em gomos para ser comido em família nos fins de semana[1]. Mas daí a afirmar que a mente não tem papel algum na produção e determinação dos vários tipos de comportamento parece ser mais difícil de sustentar. É bem verdade que protozoários não têm mentes (aliás, não têm sequer cérebros) e nunca precisaram dessas para sobreviver. Mas se compararmos os comportamentos de um protozoário com os de um ser humano encontraremos, com certeza, uma diferença razoável – uma diferença no mínimo em termos de flexibilidade, além de um repertório de ações possíveis incomparavelmente mais rico. Como explicar essas diferenças sem recorrer ao papel desempenhado por uma mente na determinação e na modificação do comportamento? E como admitir esse papel sem pressupor a existência de uma interação entre mente e corpo na produção do comportamento – uma interação que se tornaria inconcebível se sustentássemos, como quer o dualista substancial, que o mental é imaterial?

Se o dualista substancial encontra-se num beco sem saída não é o que ocorre, aparentemente, com o dualista de propriedades. Esse último tem um outro ponto de partida: a ideia de que estados mentais são uma *propriedade especial* ou um atributo específico de algumas porções da matéria da qual é composto o universo. O dualista de propriedades (*property dualist*) rejeita o fisicalismo, mas sustenta que essa rejeição não implica necessariamente postular a existência de uma substância adicional – que seria a substância mental. Essa propriedade especial *emerge* da substância material, mas, à diferença do emergentismo materialista, o dualista de propriedades sustenta que tal

1. Esta sentença jocosa é do filósofo I. Kant.

propriedade especial não pode ser descrita em termos físicos. É o próprio cérebro que produz a subjetividade e os estados subjetivos, mas esses nunca poderiam ser *integralmente* mapeados em relação a estados cerebrais. O mental supervém ao físico, mas determina algo *para além* das propriedades físicas.

O dualista de substâncias identifica a mente com uma substância imaterial ou uma quintessência ao modo cartesiano. Ora, postular uma substância imaterial adicional ao nosso corpo não *explica* como essa substância poderia gerar estados subjetivos. Sensações, por exemplo, são um tipo de estado subjetivo. Mas como explicar sua natureza simplesmente postulando a existência dessa substância imaterial? Como poderia afirmar que elas ocorrem *em mim* se essa substância imaterial não fosse parte de meu corpo? Tampouco o materialismo fisicalista pode nos explicar a natureza dos estados subjetivos: ele não nos mostra como poderíamos passar de uma perspectiva de terceira pessoa para uma perspectiva de primeira pessoa sem romper com a linguagem extensional da ciência. É desse cenário de insatisfação que o dualismo de propriedades deriva sua força. A alternativa seria supor que uma única e mesma substância, qual seja, o cérebro, pode instanciar propriedades físicas e, *além dessas*, propriedades mentais ou estados subjetivos. Essa é a aposta do dualista de propriedades.

Contrariamente ao dualismo substancial, o dualista de propriedades pode sustentar que alterações físicas no cérebro podem resultar em alterações na mente. Haveria algum tipo de conexão causal entre mente e cérebro. Contudo, o dualista de propriedades não aceita que uma descrição física do mundo possa ser tão completa a ponto de nela poderem ser incluídos também os fenômenos mentais – caso contrário o mental não seria algo *para além* das propriedades físicas. Contra o reducionista, o dualista de propriedades não supõe ser possível encontrar uma explicação para a natureza do mental pela redução deste último a um conjunto de propriedades físicas do cérebro e do sistema nervoso. Em outras palavras, o dualista de propriedades aposta na desesperança de se encontrar uma interpretação física dos estados subjetivos e conscientes.

Uma caracterização mais precisa das posições defendidas pelos dualistas de propriedades pode, entretanto, revelar algumas de suas dificuldades teóricas. Não há dúvida de que tentativas de uma abordagem reducionista da natureza do mental têm sido, até o momento, malsucedidas. Investigações detalhadas da composição física de nosso cérebro não forneceram, até agora, nenhuma pista para sabermos como desse podem surgir a subjetividade e a consciência. A análise do substrato material de nossos cerebros não tem revelado, até agora, nada para além de um conjunto de propriedades químicas e físicas que compõem outros segmentos do mundo material, igualmente composto por um conjunto de partículas subatômicas e suas interações. Nesse sentido, o dualista de propriedades estaria correto na sua crítica ao reducionismo.

Mas é precisamente essa semelhança entre nosso cérebro e outros segmentos do mundo material que nos causa perplexidade. Superar essa situação de perplexidade significa apostar na possibilidade de algum dia encontrarmos alguma característica específica na composição do cérebro que explique como dele emergem estados subjetivos e conscientes sem ter de recorrer a algum tipo de propriedade oculta ou miste-

riosa[2]. Mas essa é uma hipótese que o dualista de propriedades não pode endossar, caso contrário sua posição não poderia ser caracterizada como um tipo de dualismo, tornando-se autocontraditória.

Ora, mas ao sustentar que a produção de experiências subjetivas e conscientes não pode surgir de nenhuma característica física específica do cérebro – por estar além delas ou ser irredutível a elas – o dualista de propriedades incorre num dilema. Ou bem ele admite a existência de características específicas do cérebro que seriam responsáveis pela produção da subjetividade e da consciência – tornando sua posição autocontraditória – ou bem ele admite que qualquer elemento do mundo material poderia, em última análise, produzir uma mente. Essa segunda alternativa é a única que pode ser adotada pelo dualista de propriedades para escapar do risco da autocontradição. Adotar essa alternativa, contudo, pode levar-nos à posição denominada de pampsiquismo. O panpsiquismo abre a possibilidade de que qualquer elemento do mundo material, na medida em que não difira essencialmente dos elementos presentes nos cérebros, poderia eventualmente instanciar propriedades mentais. Tudo no mundo poderia, *em princípio*, ter uma mente ou produzir uma mente. O pampsiquismo se afigura uma teoria bizarra, pois a partir dela podemos voltar a algo parecido com uma imagem animista do mundo, onde tudo poderia ter uma mente, inclusive as pedras e os trovões. A aceitação do pampsiquismo parece, contudo, ser inevitável para o defensor do dualismo de propriedades.

Não é impossível que o pampsiquismo venha a triunfar algum dia, ou seja, que ele se consolide no esteio do sonho de uma teoria unificada da natureza – uma teoria que juntaria o físico e o mental num único tipo de substância com a propriedade de ser física e mental ao mesmo tempo. Essa é uma possibilidade perfeitamente concebível embora ela ainda nos pareça, nos dias de hoje, muito implausível. Aceitá-la parece ser o preço que temos de pagar pela adoção do dualismo de propriedades. Tudo dependerá de quanto os argumentos dos principais defensores do dualismo de propriedades no século XX, Thomas Nagel e David Chalmers, possam nos convencer de que essa seria a melhor posição filosófica a ser adotada. Em outras palavras, a aceitação do pampsiquismo como uma consequência natural do dualismo de propriedades, a despeito de sua implausibilidade, será proporcional à força dos argumentos desses filósofos e à sua capacidade de persuasão filosófica. Examinaremos esses argumentos a seguir, mas deixaremos o julgamento final a critério do leitor.

---

2. Livrar-se de propriedades ocultas ou misteriosas e conseguir uma explicação da natureza da consciência em termos neurobiológicos parece ser a esperança não apenas de neurocientistas mas também de alguns filósofos. Searle (1997), por exemplo, assinala que a investigação científica dessas propriedades possibilitará transformar o "mistério da consciência" no "problema da consciência".

Thomas Nagel

O filósofo norte-americano Thomas Nagel é um dos maiores defensores do dualismo de propriedades no século XX. Em dois artigos célebres, "Physicalism" [O Fisicalismo][3] (1965) e "What is it like to be a bat?" [O que é ser como um morcego?] (1974) ele argumenta em favor do dualismo de propriedades tomando como ponto de partida que:

> 1) Existem alguns estados mentais que, por suas propriedades específicas, não podem ser descritos a partir de um vocabulário fisicalista. O vocabulário fisicalista é essencialmente intersubjetivo. O mesmo ocorre com nossa linguagem cotidiana. Mas em ambos os casos, mesmo que possamos nos referir a esses estados usando nossa linguagem, essa apenas resvala neles, pois sua verdadeira compreensão exigiria mais do que a simples possibilidade de referir-se a eles ou de descrevê-los através da linguagem.
>
> 2) A pressuposição para a compreensão desses estados mentais através da linguagem é a existência de um ponto de vista subjetivo que seria irredutível à imagem científica do mundo proporcionada pela descrição fisicalista. Nada se assemelha ao ponto de vista subjetivo; além de irredutível, ele é único.

No seu artigo "What is it like to be a bat?" Nagel argumenta em favor desses dois pontos de vista. Seu ponto de partida é o caráter imediato da experiência consciente. Todos nós temos experiências conscientes e sabemos que as temos, sem que, entretanto, isso nos permita dizer o que é ser consciente ou o que é ter experiências conscientes. Desse caráter paradoxal da experiência consciente, Nagel deriva ainda um outro aspecto: sua intransponibilidade enquanto experiência subjetiva. Como não posso dizer o que é ser consciente ou ter experiências conscientes, não posso projetar para outros seres o que é ter experiências conscientes iguais às minhas. Não posso, tampouco, saber exatamente o que é, para outras criaturas, vivenciar suas experiências conscientes. Não posso saber *o que é* ser como um morcego, embora eu possa *imaginar* o que é ser como um morcego. Em outras palavras, não posso me transpor para a perspectiva de mundo que o morcego tem, pois isto exigiria não apenas que meu organismo fosse igual ao do morcego como também que eu pudesse ocupar a mesma perspectiva de mundo que o morcego tem. A perspectiva de mundo do morcego é intransponível para mim, da mesma forma que a minha é intransponível para o morcego.

Ser um morcego significa adotar uma perspectiva *específica* de mundo. Essa perspectiva específica é determinada não apenas pelas diferentes perspectivas espaciais que o morcego pode vir a ter – e que serão únicas na medida em que eu nunca poderia ocupar simultaneamente os espaços ocupados pelo corpo do morcego – como também pelo fato de que o aparato sensorial utilizado pelo morcego para experienciar o mundo

---

3. Há tradução desse artigo para o português, cf. Teixeira (1996b).

é totalmente diverso do meu. Os morcegos não têm experiências visuais, localizam-se e locomovem-se pelo mundo através de um sistema de radar; um sistema de ecolocalização. Através desse sistema, os morcegos identificam a presença de objetos e obstáculos detectando os reflexos sonoros de seus próprios guinchos rápidos, subitamente modulados e de alta frequência. Isto determina um modo específico não apenas de perceber o mundo como também um modo específico de vida interior cuja natureza última é indescritível por ser essencialmente subjetiva. É essa indescritibilidade que não nos permite extrapolar a nossa própria experiência para o que seria a experiência de um morcego. É por isso que nunca poderemos saber *o que é ser como um morcego*, no máximo, podemos nos imaginar imitando um morcego, mas nossa imaginação se restringe aos recursos de nossa própria mente. E, mesmo que pudéssemos gradualmente sofrer uma metamorfose que nos transformasse em morcegos não poderíamos, do nosso ponto de vista do presente, imaginar a experiência dos momentos posteriores à transformação. Nossa linguagem nunca poderia se adaptar plenamente à experiência de mundo do morcego.

Essa indescritibilidade é sintoma da intransponibilidade da experiência, ou seja, do caráter subjetivo último da experiência que, por não deixar-se descrever, não pode, tampouco, ser adequada e completamente exportada para o ponto de vista de outras criaturas. Isto significa que qualquer descrição da experiência será sempre necessariamente incompleta – ou seja, a descrição da verdadeira natureza da subjetividade é apenas resvalada pela linguagem. Encontramos aqui a ideia de que uma perspectiva dualista se justificaria na medida em que o vocabulário de uma linguagem fisicalista seria sempre incompleto. Haveria mais coisas no mundo do que aquilo que é proporcionado por sua descrição objetiva.

O passo seguinte consiste em mover-se do caráter subjetivo e intransponível da experiência para o seu caráter essencialmente *privado* – o que reforça a existência de experiências irredutíveis à linguagem e a qualquer descrição fisicalista do mundo. Imaginemos, por exemplo, que estamos estudando a fisiologia do olho de um inseto. Vários insetos têm um sistema ocular bastante distinto do nosso. Alguns deles têm olhos que parecem domos, cada um deles estando situado num dos lados da cabeça. Esses olhos são compostos de milhares de facetas ou módulos minúsculos. A luz incide sobre cada um deles, que funciona como uma poderosa lente de aumento. O grau de resolução das imagens produzidas pelos olhos da maioria dos insetos é precária; acredita-se também que eles não possam identificar cores.

Nagel diria que, por mais que estudemos a fisiologia do olho de um inseto, não poderíamos saber o que é ter suas experiências visuais: poderíamos, no máximo, imaginá-las. Podemos, contudo, estender ainda mais as consequências dessa intransponibilidade de experiência e de ponto de vista em relação ao mundo.

Imaginemos uma situação na qual alguém queira conceber algum tipo de experimento para verificar qual o espectro de cores percebido por um determinado inseto. Eu poderia expor esse inseto a diversos tipos de cores para em seguida verificar, quais dentre essas cores levam a algum tipo de reação no seu sistema ocular, ou seja, eu po-

deria tentar saber quais as cores que são discriminadas pelo inseto. Meu ponto de partida seria, contudo, o espectro de cores que é discriminado por mim, na qualidade de ser humano. Nesse sentido, eu poderia apenas verificar se o inseto discrimina as cores que eu, como ser humano, sou capaz de discriminar. Eu discrimino o violeta do azul; posso tentar examinar o sistema ocular do inseto para tentar saber se ele também reage diferencialmente ao violeta e ao azul. Mas, na verdade, sempre estarei projetando minhas próprias discriminações àquilo que seria a experiência visual colorida do inseto: nunca poderia saber se o inseto experiencia o violeta e o azul da mesma maneira que eu os experiencio, mesmo que eu constate que o inseto, da mesma maneira que eu, reage diferentemente à exposição ao violeta e ao azul. A experiência de cor do inseto será sempre inescrutável, mesmo que eu refine meu experimento ao ponto de não mais falar de azul ou de violeta, mas apenas de uma frequência de onda que corresponde ora ao azul ora ao violeta. Encontraríamos aqui elementos da experiência subjetiva de um organismo qualquer que seriam, rigorosamente falando, inescrutáveis. Em filosofia da mente designamos esses elementos subjetivos inescrutáveis pelo termo técnico *qualia*[4].

Os *qualia* apontam para a existência de elementos da experiência humana que seriam inescrutáveis e incomunicáveis mesmo entre seres humanos que partilham de uma mesma linguagem e de uma mesma perspectiva específica de mundo. Essas experiências, por serem privadas e inescrutáveis seriam apenas parcialmente descritíveis pela linguagem. Detectamos a existência desses *qualia* – privados e inescrutáveis – ao considerarmos que o caráter subjetivo de certas experiências não poderia ser completamente mapeado em relação a qualquer estado cerebral. O exame e a descrição de um estado cerebral correlato a uma determinada experiência seria insuficiente para determinar como seus *qualia* estariam sendo vivenciados pelo sujeito. Em outras palavras, os *qualia* não seriam capturados pela linguagem intersubjetiva sobre a qual se assenta a ciência.

Da mesma maneira que podemos apenas *imaginar* o que é o mundo para um morcego (ou descrevê-lo precariamente sem, entretanto, saber o que é *sentir-se* como um morcego), será que poderíamos descrever *o gosto do sal* para alguém que nunca o experimentou – como é o caso, por exemplo, dos índios à época da descoberta? Certamente qualquer descrição seria redundante e, no máximo, o que poderíamos dizer é que o gosto do sal "é salgado". Somente após experimentar o sal um índio poderia partilhar da experiência (do *qualia*) de saber o que é o gosto do sal. A descrição plena do gosto do sal pressuporia, assim, uma experiência comum partilhada por dois indivíduos da mesma espécie sem o que esse *qualia* específico permaneceria inescrutável.

Ora, esse mesmo tipo de argumento foi proposto pelo filósofo australiano F. Jackson nos seus artigos "Epiphenomenal *Qualia*" [*Qualia* epifenomênicos] (1982) e

---

4. A palavra *qualia* originou-se da ideia de que os objetos que percebemos teriam qualidades primárias (forma, tamanho, solidez etc.) e qualidades secundárias (cores, aromas etc.). *Qualia* designa as qualidades secundárias e a ideia de que estas só existem em nossas mentes.

"What Mary didn't know" [O que Mary não sabia] (1986). Nesses artigos, Jackson conta a história de uma brilhante neurocientista chamada Mary, a qual tem um defeito de nascença em sua visão que a impede de perceber cores: ela só percebe o mundo em branco e preto. Esse defeito não a impediu, contudo, de estudar e tornar-se uma cientista notável. Mary sabe tudo acerca da visão humana, ou seja, como os olhos percebem o mundo, como esses dados são registrados pela retina e processados pelo cérebro. Ela sabe tudo sobre a natureza das cores – tudo o que a ciência pode descrever acerca das cores e como essas são percebidas pelos humanos. Em outras palavras, Mary tem uma descrição física completa das cores e de como elas são percebidas. Essa descrição seria, contudo, insuficiente para Mary saber *o que é* ter uma experiência do vermelho.

Um dia Mary submete-se a uma neurocirurgia bem-sucedida e passa a enxergar o mundo em cores. Nesse mesmo dia, Mary, colocada diante de um tomate maduro, tem uma primeira experiência do vermelho. Ela passa a saber o que é experienciar alguma coisa como sendo vermelha. Desse argumento (ou dessa alegoria) Jackson tira a conclusão de que as experiências subjetivas são algo mais do que um conjunto de condições físicas que as proporcionam. Essas experiências subjetivas se sobrepõem a qualquer tipo de descrição física que possamos ter do nosso funcionamento cerebral, são sempre algo mais do que uma descrição completa do modo como as cores são processadas pelo cérebro. Experiências subjetivas apontariam para a existência de algo não físico e inescrutável.

Essa é sem dúvida uma posição tentadora, mas será que faria sentido estendê-la para o âmbito de todas as nossas experiências? Por que não poderíamos pensar ao contrário e imaginar, por exemplo, que se construirmos um robô dotado de um cérebro igual ao nosso, quando esse robô detectar algo salgado terá um estado cerebral igual ao que ocorre em nosso cérebro e, portanto, uma experiência de provar algo salgado? O que nos impediria de construir essa máquina? Churchland (1995) objeta ao argumento de Jackson afirmando que do caráter subjetivo da experiência não podemos *necessariamente* concluir que sensações e percepção de cores devam ser consideradas algo não físico. Esse seria um passo indevido que estaria sendo dado por filósofos como Nagel e Jackson, pois nada impediria que algum dia a neurociência viesse a descobrir exatamente quais são os correlatos cerebrais das experiências subjetivas. Churchland não concorda com o caráter necessariamente inescrutável que as experiências subjetivas teriam – que razões teríamos para supor que a neurociência um dia não possa desvendar os correlatos neurais da subjetividade?

O mesmo tipo de contra-argumento encontramos em Dennett (1991). Dennett concorda com a existência de estados subjetivos, mas não com sua inescrutabilidade. Para ele, supor que estados subjetivos impliquem inescrutabilidade equivaleria a adotar o ponto de vista cartesiano, ou seja, a ideia de que deveríamos supor a existência de algum tipo de quintessência subjacente a nossos estados mentais – uma quintessência que da qual se derivaria necessariamente a conclusão de que a subjetividade e a consciência são irreplicáveis. Mas por que teríamos que pressupor a existência dessa quintessência, esse artigo de fé da metafísica cartesiana?

Se todas as experiências forem subjetivamente intransponíveis e determinadas por um ponto de vista único, um médico seria incapaz de diagnosticar uma hepatite sem ter sofrido anteriormente desse tipo de doença. A ginecologia, por exemplo, teria de se tornar uma profissão exclusivamente feminina – o que certamente seria um contrassenso. Se, no limite, todas as experiências fossem subjetivas e privadas, a própria linguagem como instrumento de comunicação tornar-se-ia impossível. Chegaríamos ao contrassenso de ter uma linguagem que seria de uso exclusivo, com a qual nos referiríamos unicamente a nossas próprias sensações e a estados mentais privados. Supor que essa linguagem poderia existir equivaleria a tornar-nos praticamente esquizofrênicos, falando uma linguagem cujo objetivo não seria a comunicação; uma verdadeira contradição em termos.

Ora, até que ponto poderíamos estender a ideia de que um ponto de vista único em relação ao mundo – como aquele que certamente cada um de nós tem – poderia forçar-nos a abandonar uma perspectiva fisicalista? A ideia de uma perspectiva única não determina, mas certamente contribui para a formação da noção de "eu" ou de *self* – uma ideia que, segundo Nagel, torna-se inconcebível se adotarmos uma perspectiva fisicalista. Esse é o tema de seu artigo "Physicalism" [O fisicalismo] publicado em 1965.

Nele, Nagel argumenta que qualquer concepção de identidade pessoal torna-se necessariamente incompatível com o fisicalismo. Se à proposição "Eu sou o JFT" correspondesse um estado físico ou estado cerebral, seria possível que esse estado físico ou estado cerebral ocorresse também em outra pessoa. Embora eu pudesse admitir que essa pessoa poderia ter a mesma perspectiva de mundo que eu tenho, não faria sentido supor que ela *também* é o JFT. Haveria, assim, *pelo menos um* estado mental ao qual não poderia corresponder um estado cerebral, o que já seria suficiente para solapar a teoria da identidade mente-cérebro.

Ora, se por um lado pode ser difícil ter de abrir mão do fisicalismo, por outro lado precisamos considerar que é igualmente difícil abrir mão de uma concepção de "eu" ou de *self*. Certamente não conseguiremos encontrar nenhuma ideia correspondente ao "eu" ou ao *self* através de qualquer ato de introspecção. Mas será possível abrir mão da ideia de que meus estados mentais formam algum tipo de unidade – por mais imperfeita e frágil que essa seja? Ora, como eu poderia ligar esses estados mentais uns aos outros? Conceber que essa ligação seria através de algum tipo de conexão física entre os vários estados de meu cérebro seria insuficiente. Pois eu poderia conceber, na qualidade de um experimento mental, que, através de uma cuidadosa cirurgia alguém conseguisse conectar fisicamente meu cérebro com o cérebro de outra pessoa. Poderíamos imaginar que, através de nervos artificiais podemos ligar o cérebro de uma pessoa a outra. Essa ligação formaria, então, algo como uma terceira pessoa, resultante da ligação de meus estados mentais com os estados mentais de uma segunda pessoa. Contudo, se essa terceira pessoa resultante dessa ligação física começasse a atuar no mundo, suas ações seriam ininteligíveis, pois não seriam ações resultantes de nenhuma das duas pessoas que tiveram seus cérebros fisicamente conectados – não seriam ações re-

sultantes de minhas crenças e desejos, nem tampouco resultantes das crenças e desejos dessa segunda pessoa com a qual fui fisicamente conectado. Ou seja, a conexão física seria insuficiente para produzir qualquer tipo de unidade de estados mentais, qualquer tipo de *self*. A ideia de "eu" e de *self* requer algo para além ou essencialmente diferente de uma unidade física. Essa seria a situação de dois irmãos siameses que nascessem partilhando um mesmo cérebro. Seriam suas experiências descritas como pertencentes a um mesmo *self* ou formando dois *selves* distintos?

Esses argumentos de Nagel podem impressionar, à primeira vista, qualquer leitor e induzir ao abandono do fisicalismo. Contudo, antes que o leitor comece a tirar suas próprias conclusões, seria melhor considerar dois contra-argumentos às concepções de Nagel. Em primeiro lugar, vale considerar que não há assunto sobre o qual se tenha falado mais na literatura de filosofia da mente, nos últimos anos, do que os *qualia*. Toda essa literatura, parece, contudo, basear-se na herança cartesiana que estipula a existência de uma distinção nítida entre propriedades primárias e propriedades secundárias de objetos físicos e como essas se refletem nas nossas sensações. Descartes sustentou que a extensão (a *res extensa*) é a propriedade primária por excelência, intrínseca aos objetos físicos que percebemos. Sustentou também que aromas, odores etc., são propriedades secundárias e foi seguido por outros filósofos como J. Locke (1632-1704) que mantiveram essa distinção. Mas será essa distinção tão nítida assim? Não seria contraintuitivo dizer, por exemplo, de um pedaço de queijo gorgonzola que seu cheiro é apenas uma qualidade secundária?

A segunda objeção diz respeito aos argumentos de Nagel acerca da impossibilidade de mapear o *self* a algum tipo de correlato cerebral. Nagel não considera a possibilidade de que a unidade da experiência ou da consciência, acessíveis apenas introspectivamente (de onde se originam nossas ideias de *self*), possam ser apenas uma ilusão bem montada. Falaremos disto no próximo capítulo.

David Chalmers[5]

Seria possível ser dualista sem ao mesmo tempo abraçar qualquer tipo de compromisso religioso? E seria possível ser dualista e, ainda assim, compatibilizar essa posição com a existência futura de máquinas inteligentes, isto é, sem romper com o programa teórico da inteligência artificial?

Esse é o tipo de dualismo desenvolvido pelo filósofo australiano David Chalmers. Seu livro *The Conscious Mind* [A mente consciente], publicado em 1996, constitui uma das tentativas mais recentes de formular uma teoria abrangente da natureza da mente e da consciência. Sua visão é ousada e caminha na direção oposta de tudo o que os cientistas cognitivos e neurocientistas desejam: reduzir estados conscientes a uma possível base neurofisiológica ou física.

---

5. Algumas passagens desta seção foram adaptadas de Teixeira (1997).

Chalmers toma como ponto de partida aquilo que para muitos – aí incluídos até alguns neurocientistas[6] – constitui o horizonte intransponível de qualquer teoria científica da natureza da consciência: reconhecer que não é possível formular uma teoria que explique plenamente como um sinal cerebral pode dar origem a um estado consciente. Ele sugere que uma teoria da consciência deve tomar a noção de *experiência consciente* como sendo um elemento básico. A experiência consciente deve ser considerada como uma característica fundamental do mundo, do mesmo jeito que massa, carga eletromagnética e espaço-tempo. Muitos fenômenos são explicáveis em termos de entidades mais simples do que eles, mas isto não pode ser generalizado. Às vezes, certas entidades precisam ser tomadas como primitivas ou *fundamentais*, ou seja, como constituindo entidades cuja natureza não pode ser explicada em termos de algo mais simples. Por exemplo, no século XIX ficou claro que processos eletromagnéticos não poderiam ser explicados em termos de processos mecânicos, e Maxwell introduziu as noções de carga e força eletromagnética como componentes fundamentais de sua teoria física. Ou seja, para explicar o eletromagnetismo, a ontologia da física teve de ser expandida. Outras características que a teoria física assume como fundamentais são as noções de massa e de espaço-tempo. Nunca se procurou explicar essas noções em termos de algo mais simples, o que entretanto não descarta a possibilidade de se construir uma teoria da massa ou do espaço-tempo.

Chalmers sustenta que consciência e experiência subjetiva devem ser tomadas como elementos básicos ou fundamentais de qualquer teoria da mente; que essas devem ser ponto de partida e não ponto de chegada, por não serem passíveis de uma redução ou explicação em termos de entidades mais simples derivadas da neurociência ou da física. Essa posição é uma variedade de dualismo, na medida em que rejeita o materialismo reducionista e o fisicalismo. Mas trata-se de uma variedade "inocente" de dualismo, inteiramente compatível com uma visão científica do mundo. Como assevera Chalmers, não há nada místico ou espiritual nessa teoria. É uma teoria inteiramente naturalista, na medida em que, segundo ela, o universo não é nada mais do que uma rede de entidades básicas que obedecem a um conjunto de leis, a partir das quais podemos montar teorias, inclusive uma teoria da consciência. Trata-se de um *dualismo naturalista*.

Iniciar uma teoria tomando como elemento básico a experiência consciente, significa, antes de mais nada, reconhecer que o problema da consciência não é um pseudoproblema e que o filósofo da mente não pode fugir da tarefa de ter de enfrentá-lo seriamente. Essa tentação – fugir ou simplesmente escamotear o problema da natureza da consciência – pode surgir pelo fato de estarmos enfrentando um problema extremamente árduo. Para começar, a filosofia da mente não reconhece a existência de apenas um problema da consciência. "Consciência" é um termo polissêmico e por vezes ambíguo, que se refere a vários tipos de fenômenos, como por exemplo:

---

6. A referência é a Eccles, cf. Popper e Eccles (1977).

– a habilidade para discriminar, categorizar e reagir a estímulos ambientais;

– a integração da informação através de um sistema cognitivo;

– a capacidade de relatar a ocorrência de estados mentais;

– a habilidade de um sistema para acessar seus próprios estados internos;

– o foco da atenção;

– o controle deliberado do comportamento;

– a diferença entre sono e vigília[7].

Todos esses fenômenos estão associados com a noção de consciência. Por exemplo, diz-se que um estado mental é consciente quando é passível de ser relatado verbalmente ou quando é internamente acessível. Às vezes, diz-se que um sistema está consciente de uma informação quando tem a habilidade de reagir com base nela ou quando a integra e a elabora para produzir determinados comportamentos. Dizemos frequentemente que uma ação é consciente porque é deliberada. Outras vezes, referimo-nos a um organismo como estando consciente quando está em vigília.

No entender de Chalmers, nenhum desses fenômenos nem tampouco seu conjunto caracteriza o verdadeiro problema da consciência: eles constituem apenas os aspectos *funcionais* da experiência consciente. Isto significa dizer que, em última análise, esses fenômenos *podem vir a ser* explicados cientificamente. Em outras palavras, nada impede que algum dia eles possam vir a ser explicados seja através de um modelo computacional seja através da descoberta de mecanismos neurais. Por exemplo, para explicar o acesso e a capacidade de relatar a ocorrência de estados mentais basta especificar o mecanismo através do qual a informação acerca de estados mentais é recuperada e tornada disponível para relato verbal. Para explicar a integração da informação precisamos apenas conceber mecanismos através dos quais essa seja combinada e em seguida utilizada em outros processos. Para explicar a distinção entre sono e vigília, uma explicação, em termos neurofisiológicos, que dê conta da diferença de comportamento do organismo nesses dois estados é mais do que suficiente.

Se explicar a consciência se resumisse à explicação desses fenômenos, então não haveria um *problema filosófico* da consciência. Embora esses sejam problemas empíricos de difícil solução, eles ainda não caracterizam os verdadeiros problemas colocados pela consciência. São, em última análise, os *easy problems* [problemas simples]. A grande dificuldade é o chamado *problema da experiência* (*hard problem*). Quando pensamos e percebemos, existe um tipo de processamento de informação, mas também um aspecto subjetivo envolvido – e se não pudermos explicar a natureza desse aspecto subjetivo nunca entenderemos *o que é* a consciência. Saber *o que é* a consciência significa deslindar o "*hard problem*".

7. Cf. Chalmers (1995). A exposição abaixo segue o percurso desse artigo.

O reconhecimento da existência de um "*hard problem*" tem como consequência uma desqualificação das tentativas de explicação funcional da natureza da consciência entendida como *experiência consciente*. Explicações funcionais podem ser necessárias, mas certamente não serão suficientes para explicar a natureza da experiência consciente. Pois, como explicamos o desempenho de uma função? Especificando *o mecanismo que desempenha a função*. A aplicação de conhecimentos oriundos da neurofisiologia e das ciências cognitivas pode resolver vários problemas nesse sentido. Se mostrarmos como um mecanismo neuronal ou computacional pode desempenhar uma determinada tarefa, teremos explicado o fenômeno em questão.

Mas no caso da *experiência consciente* esse tipo de explicação falha. O problema da experiência consciente requer algo mais do que explicar o desempenho de funções. Em outras palavras, o *hard problem* persiste mesmo quando o desempenho de todas as funções relevantes é explicado. A questão que se coloca é a seguinte: *Por que o desempenho dessas funções no cérebro é acompanhado por experiências?* Ou seja, pode-se explicar como a informação é discriminada, integrada e relatada, mas isto não significa explicar como ela é *experienciada*. Essa é a questão-chave no problema da consciência – explicar como e por que surge a experiência no decorrer do processamento de informação. Não existe nenhuma função cognitiva cuja explicação leve automaticamente a uma explicação da experiência consciente. A experiência consciente *supervém* à sua base física, ou seja, nenhum fato do mundo, mesmo em nível microfísico, implica necessariamente na produção de estados conscientes.

O conceito de *superveniência*, cuidadosamente analisado por Chalmers em seu livro, sustenta esse ponto de vista. Uma propriedade *B* de um determinado indivíduo é chamada de superveniente se é produzida por um conjunto de propriedades *A* desse mesmo indivíduo. Por exemplo, um conjunto de propriedades físicas pode determinar um conjunto de propriedades biológicas na medida em que fenômenos vitais dependem de uma base física. Esses fenômenos vitais são, então, *supervenientes* em relação à sua base física; se as propriedades físicas variarem, as propriedades biológicas também variarão. A determinação de propriedades supervenientes pode ser *lógica* (conceitual) ou *natural* (empírica ou nômica, isto é, decorrente de uma lei da natureza). No caso da superveniência lógica as propriedades *B* são consequência automática da existência das propriedades *A*, ou seja, não seria possível conceber *A* sem conceber *B*. Já no caso da superveniência natural é possível conceber *A* sem conceber *B*, mas existe uma conexão empírica, *de fato*, entre *A* e *B*.

Ora, o esforço de Chalmers será mostrar que estados conscientes não são logicamente supervenientes em relação a estados físicos: é perfeitamente concebível a existência de duas criaturas fisicamente idênticas, sendo que uma desenvolve experiências conscientes e outra não. O exemplo paradigmático invocado por Chalmers é a plausibilidade de concebermos criaturas como zumbis. Nesse experimento mental, um zumbi é uma criatura fisicamente idêntica a mim, molécula por molécula. Ele é também funcionalmente equivalente a mim, no sentido de que pode fazer tudo o que eu faço. Contudo, posso perfeitamente conceber que esse zumbi não tenha experiências

conscientes. Esse zumbi pode até ser uma réplica de mim mesmo, mas replicar minhas características físicas e funcionais não implica, automaticamente, replicar minha possibilidade de ter estados conscientes. O mesmo poderia ser dito de um robô que replicasse totalmente minhas possibilidades funcionais, um robô humanoide como é o caso do COG[8]. Assim sendo, nada indica que estados conscientes sejam necessariamente supervenientes em relação a estados físicos e nem mesmo a determinadas arquiteturas funcionais. Estados conscientes são, no máximo natural ou empiricamente supervenientes em relação a estados físicos, ou seja, não há conexão lógica entre base física ou arquitetura funcional e consciência. A consciência é contingente em relação a sua base física, ela é um *fator suplementar*. A experiência consciente *pode* emergir de uma estrutura física, mas não é consequência necessária dessa, isto é, não *deriva* dela.

Chegamos assim à proposta de uma teoria não reducionista da experiência consciente. Essa teoria é um "dualismo brando" ou "dualismo metodológico" que deve especificar um conjunto de princípios básicos que nos mostrem como a experiência consciente *supervém* às características físicas do mundo. A ideia do "dualismo metodológico" é que nem todos os sistemas físicos têm características sobre as quais a experiência consciente possa supervir. Essas características estão presentes em nosso cérebro; nele a experiência consciente supervém e é um dado imediato – sabemos que somos seres conscientes. Um outro sistema físico qualquer, um robô, por exemplo, pode instanciar essas características físicas de outra maneira, pois essas são, na verdade, um conjunto de princípios – *princípios psicofísicos*. Preencher esse requisito, contudo, não significa que o robô vá ter experiências conscientes, mas apenas que ele se torna um candidato a tê-las – um candidato cujo sucesso é bastante improvável e que possivelmente nunca chegará a ser mais do que um zumbi. Ou seja, os princípios psicofísicos estabelecem apenas as condições necessárias para um paralelismo entre mente (ou consciência) e cérebro.

Chalmers identifica três princípios psicofísicos na sua teoria: o princípio de coerência estrutural, o princípio de invariância organizacional e o princípio do duplo aspecto da teoria da informação. O primeiro princípio estabelece uma relação coerente entre a *structure of consciousness* e a *structure of awareness*[9]: toda experiência consciente é cognitivamente representada, ou seja, assume a forma de um processo cognitivo, embora nem tudo o que seja cognitivamente representável seja necessariamente consciente. Existe uma relação íntima entre cognição e consciência que torna os estados conscientes passíveis de relato verbal, acessíveis aos sistemas centrais que controlam o comportamento e tudo o mais que compõe a *structure of awareness*. Esse quase

8. COG é o nome de um robô que teve projeto desenvolvido no laboratório de inteligência artificial do MIT, pela equipe de R. Brooks. O projeto prevê a construção de um robô humanoide, uma máquina geral que possa fazer tudo o que um ser humano faz.

9. Estes são termos de difícil tradução para a língua portuguesa. *Consciousness* significa "estar consciente de" e *awareness* significa "estar ciente de". A língua inglesa estabelece essa distinção sutil; em português, tenderíamos a traduzir "consciousness" e "awareness" por "consciência", o que não captaria essa diferença.

isomorfismo entre *structure of consciousness* e *structure of awareness* permite que teorias cognitivas e neurofisiológicas sirvam de ponto de partida para uma teoria da experiência consciente: essas teorias devem explicar a base física ou os correlatos neurofisiológicos sobre os quais a experiência consciente supervém.

O princípio da invariância organizacional estipula que dois sistemas com a mesma organização funcional poderão ter experiências qualitativamente idênticas. Isto significa dizer que se construirmos uma réplica do cérebro humano em silicone preservando, contudo, os mesmos padrões causais de organização neuronal, esse cérebro replicado poderá ter as mesmas experiências que o cérebro humano. O que conta na emergência de experiências não é o tipo de substrato físico de um sistema, mas seu princípio arquitetônico ou a organização de seus componentes. Padrões organizacionais, instanciados em diferentes tipos de substrato são necessários, mas não suficientes, para o surgimento da experiência consciente.

O terceiro princípio, do duplo aspecto da informação, é o princípio básico e fundamental da teoria da consciência de Chalmers. Ele toma como ponto de partida a noção de informação tal como é definida por Shannon (1948) e sustenta que essa tem um duplo aspecto: um físico e outro fenomênico. É o aspecto fenomênico que dá origem à experiência consciente e esse princípio é, sem dúvida, o mais controverso na teoria de Chalmers: afinal, quais são as peculiaridades da informação que podem dar origem a estados conscientes? Será a consciência privilégio apenas de cérebros humanos – onde esse duplo aspecto está presente – ou poderá ela ser estendida a outros processadores de informação como cérebros de animais ou até mesmo máquinas?

É notável o quanto este aspecto permanece obscuro na teoria de Chalmers e o situa ao lado do grupo de filósofos contemporâneos como McGinn (1991) que foram chamados de *new mysterians* [os "novos misterianos"] por suporem que há algo intrinsecamente misterioso, intransponível, que nos impede de chegar a uma explicação completa da natureza da consciência[10]. Em várias passagens de seu livro nota-se um constante flerte com posições dualistas estritas que são, em seguida, abrandadas pela ideia de um "dualismo naturalista"[11]. Afinal, ao reconhecer que a "experiência consciente" é uma dimensão qualitativa do universo ou um "primitivo" da mobília do mundo estaremos tão distantes assim da ideia cartesiana da pluralidade das substâncias? Pouco podemos dizer do "aspecto dual da informação" da mesma maneira que pouco se pode dizer das características da "substância pensante" cartesiana. A irredutibilidade da dimensão subjetiva da experiência consciente parece originar-se do fato dessa apresentar-se como um dado imediato –, mas será esse o único ponto de partida plausível para iniciarmos uma teoria da consciência? Por que teríamos de necessariamente iniciar

10. Numa entrevista concedida a Robert Wright, da revista *Time* de abril de 1996, McGinn afirmou que seria tão difícil para os seres humanos resolver o problema da consciência quanto para uma lesma compreender teoria psicanalítica.

11. A estranheza dessa posição, o "dualismo naturalista", é reconhecida pelo próprio Chalmers. Cf., por exemplo, as passagens em Chalmers (1996, p. 357s.).

nossa reflexão assumindo uma posição solitária? Não será a intransponibilidade entre primeira e terceira pessoa – tão cara e sempre relembrada pelos filósofos dualistas – uma consequência inevitável que surge de tomarmos sempre essa posição ao iniciarmos nossas tentativas de explicar a natureza da consciência? Se tomamos como ponto de partida nossa experiência consciente, nunca poderemos estar seguros de poder atribuí-la a outras criaturas. Mas não será isto contraintuitivo? Quando olhamos para uma lagosta sendo jogada na água quente, contorcendo-se com a dor, não estamos intuitivamente atribuindo algum tipo de experiência consciente a esse organismo? Ou existirá dor sem consciência?

O flerte de Chalmers com o cartesianismo torna-se igualmente evidente na sua teoria da superveniência dos estados conscientes. A crítica a explicações reducionistas e puramente funcionais da natureza da consciência encontra-se, de maneira embrionária, nos escritos de Descartes sobre os autômatos. Descartes sustentava (cf. o capítulo II) que a duplicação de características materiais e funcionais de um ser humano poderia ser condição necessária, mas não suficiente para se replicar a vida mental humana[12]. Um autômato bem construído pode vir a fazer tudo o que um ser humano faz, mas nunca se igualaria a esse: seria, no máximo, uma proeza de engenharia, algo que, contudo, não teria *alma* (e não poderíamos substituir essa palavra por "experiência consciente"?). Nesse sentido, o autômato de Descartes não é muito diferente do zumbi de Chalmers. Basta que na sentença anterior troquemos a palavra "alma" pela palavra "experiência consciente" e que não afirmemos com Descartes que essa é necessariamente imaterial, sendo apenas um elemento básico do universo, que chegaremos ao "dualismo naturalista" de Chalmers.

A diferença entre a posição de Chalmers e a posição cartesiana consiste no fato de Descartes ter afirmado, categoricamente, que a vida mental *não pode supervir* no autômato. Chalmers deixa aberta essa possibilidade ao defender a inteligência artificial em sentido forte nos últimos capítulos de seu livro, embora ele talvez preferisse afirmar que autômatos serão, no máximo, zumbis. Mas a pressuposição de Chalmers de que a similaridade funcional não é suficiente e não *implica* na produção de estados conscientes é inteiramente metafísica. Afinal, se mantivermos o primado da primeira pessoa para fundar nossa teoria da consciência (como quer Chalmers), o que pode nos garantir que um robô que faça tudo o que um ser humano pode fazer não tem experiências conscientes – um robô a cujos estados conscientes não teríamos acesso?

Esta última questão faz-nos refletir sobre outros problemas que surgem a partir da teoria de Chalmers – problemas tão interessantes quanto complexos. Em primeiro lugar, destaca-se o chamado *problema da repredicação*[13]. Suponhamos que por um certo período de tempo tenhamos convivido com um robô de forma humanoide, uma réplica

---

12. Cf., por exemplo, a carta de Descartes ao Marquês de Newcastle, de 23 de novembro de 1646, onde esta posição é sustentada explicitamente. Cf. novamente a nota 7, capítulo II.

13. A este respeito cf. Gunderson (1971) capítulo I.

cuja aparência externa fosse exatamente igual à de um ser humano. Esse robô poderia ser, por exemplo, o COG, o robô humanoide desenvolvido no MIT. O COG estaria convivendo conosco e seu comportamento seria indistinguível daquele exibido por um ser humano qualquer. Ocorre que *não sabíamos* que estávamos lidando com um robô e não com um ser humano. Isto significa que por muito tempo estaríamos atribuindo ao COG os mesmos predicados mentais que normalmente atribuímos a um ser humano, incluindo a capacidade de desenvolver comportamentos e experiências conscientes. Um dia, o COG (que não sabíamos ser um robô) escorrega, cai e bate a cabeça na banheira. Seu crânio se rompe e, em vez de encontrarmos dentro dele a massa encefálica de um ser humano, encontramos fios e chips de computador. Teria cabimento *retirar* todos os predicados mentais que vínhamos atribuindo a ele até então – predicados mentais que o equiparavam a um ser humano normal? Ou teria cabimento afirmar: "bem, agora que eu descobri que você é na verdade um robô, então você não tinha estados mentais nem tampouco experiências conscientes!" Suspendemos temporariamente a atribuição de estados conscientes a uma pessoa quando essa está dormindo. Interrompemos definitivamente essa atribuição se essa pessoa entrar em coma profundo. Mas não faria sentido interromper a atribuição de experiências conscientes a essa pessoa se descobrirmos que seu cérebro é diferente do nosso (por exemplo, é de silício), sobretudo se ela continuar a agir "normalmente", ou seja, se suas ações forem indistinguíveis daquelas de um ser humano, apesar da diferença de constituição "cerebral".

A segunda questão surge no mesmo esteio da primeira: COG seria, no máximo, um zumbi. Mas será possível supor a existência de zumbis, mesmo como apenas uma possibilidade metafísica? Pois se um zumbi é, do ponto de vista comportamental, indistinguível de um ser humano, por que não atribuir a ele *também* a propriedade de ter consciência? Zumbis agem, conversam, sentem dores e passam no teste de Turing[14].

Uma terceira série de questões surge ao refletirmos sobre a noção de superveniência introduzida por Chalmers. Terá sentido, afinal de contas, afirmar que a consciência constitui um ingrediente suplementar que supervém à organização mental e funcional de um organismo ou sistema? Não estaríamos aqui diante de uma confusão conceitual? Até que ponto é sustentável a independência da experiência consciente em relação à organização funcional ou à estrutura física de um organismo? Tomemos os predicados *ser consciente* e *ter saúde*. Em ambos os casos, a atribuição desses predicados não dependeria da possibilidade de explicar o funcionamento de uma estrutura física específica de um organismo, isto é, em ambos os casos, a atribuição desses predicados fundamenta-se na observação de uma característica global do organismo. Contudo, aqui

14. O Teste de Turing, criado pelo matemático inglês homônimo, consiste em comparar os comportamentos manifestos de um organismo humano com aqueles produzidos por um robô ou computador criado para desenvolver tarefas humanas. Se da comparação resultar que as características dos comportamentos do organismo são indistinguíveis daquelas dos outputs produzidos pela máquina, podemos, de acordo com Turing, atribuir a esta estados mentais.

corremos o risco de deslizar da ideia de *característica global* para a ideia de *característica adicional*. Não teria cabimento supor que – mesmo por um ato de imaginação filosófica – poderíamos remover a saúde de um organismo ao mesmo tempo que mantemos a totalidade de seus órgãos e suas interações em perfeito estado, ou, inversamente, que poderíamos remover alguns desses órgãos e, mesmo assim, achar que preservamos a saúde do organismo ou que ela poderia permanecer intacta. Ora, por que não poderíamos afirmar o mesmo em relação à consciência?[15]

O futuro do dualismo

É difícil dizer alguma coisa acerca do futuro do dualismo. Além dos autores cujas teorias examinamos brevemente – Nagel e Chalmers –, o dualismo conta com adeptos ilustres no século XX, como, por exemplo, o filósofo da ciência Karl Popper (1902-1994) e o neurocientista John Eccles que juntaram seus esforços para defender essa posição no livro *The Self and its Brain* [O eu e seu cérebro] (1977). Contudo, conforme ressaltamos no capítulo I, o dualismo é uma filosofia sem agenda. Tudo que o filósofo dualista pode fazer é tentar convencer-nos de que mente e cérebro ou mente e matéria são radicalmente distintos e têm propriedades incompatíveis. Mas isto significa também abdicar de construir uma ciência da mente, na medida em que essa estaria fora do alcance de qualquer tipo de investigação científica.

Embora rejeitado pela ciência contemporânea, o dualismo ainda parece ser o horizonte de nossa cultura. Para percebermos o quanto ele impregna nossa vida, basta ver o quanto ele se reflete e se manifesta sutilmente na nossa linguagem – a linguagem cotidiana, que serve de base para boa parte de nossa psicologia popular. Transladamos e enterramos corpos – corpos apenas, sem suas mentes. E rezam-se missas por intenção das almas que estariam nesses corpos.

O QUE LER

CHALMERS, D. *The Conscious Mind*

NAGEL, Th. *The View from Nowhere*

POPPER, K. & ECCLES, J. *The Self and its Brain*

ROBINSON, H. *Objections to Physicalism*

15. Exploro esse argumento em Teixeira (1997), inspirado em Dennett (1995a).

CAPÍTULO V

DESFAZENDO A IDEIA DE MENTE

Uma estratégia para contornar as dificuldades colocadas pelo problema das relações entre mente e cérebro consiste em tentar desfazer nosso conceito habitual de mente, mostrando que esse se origina de algum tipo de ilusão conceitual ou linguística. Essa é uma estratégia que pode parecer, à primeira vista, bizarra, uma vez que corre em direção contrária a nossas intuições cotidianas. Sua vantagem estaria em nos livrarmos de um dos termos da equação que compõe esse tipo de problema filosófico, mostrando, ao mesmo tempo, que esse é, na verdade, um pseudoproblema.

O filósofo norte-americano Wilfrid Sellars, por exemplo, sugeriu, na década de 1960, que a ideia de mente resulta de uma espécie de ilusão cultural. No seu livro *Empiricism and the Philosophy of Mind* [Empirismo e a filosofia da mente], publicado em 1963, ele nos conta, em forma de alegoria, como teria surgido o conceito de mente.

Em tempos primordiais, numa comunidade mítica, teria aparecido um indivíduo chamado Jones – um homem com temperamento filosófico inato e grande capacidade de reflexão. Jones teria inventado o conceito de mente através de um treinamento sistemático que ele teria imposto aos membros dessa comunidade.

Jones começou por observar os comportamentos verbais de seus companheiros. De início, notou que todas as frases e sentenças usadas pelos seus companheiros se referiam apenas a coisas e eventos públicos ou seja, observáveis por todos. Mas, como bom filósofo, Jones resolveu que seria bom expandir e enriquecer essa linguagem para que ela pudesse se tornar um instrumento eficaz para identificar seres pensantes no mundo – ou seja, que essa linguagem pudesse nos distinguir como seres dotados de pensamentos, intenções, desejos e sensações. O primeiro passo dado por Jones foi criar a semântica. Nossos ancestrais míticos passaram então a caracterizar seus comportamentos verbais a partir de uma perspectiva semântica, introduzindo a ideia não apenas de que as sentenças que eles proferiam deviam ter um *significado* como também a

de que elas poderiam ser ou verdadeiras ou falsas. A ideia de significado, introduzida por Jones, marcou uma distinção inicial entre linguagem e pensamento, pois proferir sentenças deixa de ser um comportamento para tornar-se a tarefa de *expressar* o que essas pessoas estariam pensando – aquilo que se passava nas suas cabeças e que não poderia ser observado diretamente. A expansão da linguagem, através da invenção da semântica, teria sido o passo preliminar para se postular a existência de algum tipo de entidade não observável.

Ao se admitir a existência de entidades não observáveis dá-se mais um passo nesse processo de enriquecimento da linguagem empreendido por Jones. Um passo mais sutil, pois a partir daí os membros dessa comunidade mítica puderam formular a hipótese de que talvez o comportamento linguístico observável fosse causado por essas entidades não observáveis que seriam os pensamentos. A linguagem passa então a poder comportar teorias, construídas a partir dessas entidades nãoobserváveis: teorias acerca do mundo e acerca do comportamento. A expansão da linguagem teria engendrado uma expansão da ontologia dessas criaturas primevas. Haveria mais coisas no mundo do que aquilo que seria publicamente observável.

Jones pôde então conjeturar que o comportamento de seus conterrâneos poderia ser guiado por essas entidades não observáveis – os pensamentos – e que esses poderiam ocorrer mesmo quando nenhum tipo de sentença estivesse sendo proferida. Seus companheiros poderiam então "pensar" sem que "pensamento" implicasse em algum tipo de manifestação verbal ou comportamental. O intervalo entre uma ação e outra, o silêncio entre uma palavra e outra, passaram a ser vistos como o estágio preliminar onde estariam ocorrendo processos internos (não observáveis) na cabeça das pessoas – processos que culminariam com a produção de um comportamento ou de uma sentença. Esses processos internos ou pensamentos, porém, assumiriam o formato da linguagem: pensar seria produzir um discurso interno, silencioso. A expansão da linguagem engendra então a possibilidade de que Jones possa formular uma *teoria* acerca do comportamento de seus companheiros – uma teoria baseada em processos não observáveis. Mas o que poderia resultar dessa teoria? Dela resultou um novo modo de descrever o comportamento – sobretudo *nosso próprio* comportamento.

Essa história se inicia num momento no qual ainda não há linguagem e nem qualquer teoria acerca do comportamento das pessoas. Ou seja, quando não há ainda nenhum outro método de saber o que os outros estariam pensando a não ser a partir da observação de seus comportamentos. O indivíduo A observa o indivíduo B e olhando seu comportamento infere que "B está pensando em p". O indivíduo B, se olhasse para seu próprio comportamento e verificasse que esse era idêntico ao comportamento de A faria o mesmo tipo de inferência, concluindo: "Estou também pensando em p". Inicialmente, o modo como B descrevia seus processos internos, baseava-se numa comparação entre o comportamento de seus companheiros e seus próprios comportamentos. Uma comparação na qual B projetava para si mesmo o que via ocorrer com os outros.

Mas B pôde ser treinado ao ponto de poder dizer no que ele estava pensando *sem ter de observar seu próprio comportamento*. Nesse momento ocorre uma mudança ra-

dical: à medida que nossos ancestrais começaram a poder dizer no que estavam pensando sem ter de observar seu próprio comportamento, surge a ideia (ou a teoria) de que nós temos um acesso privilegiado aos nossos próprios pensamentos. O que nós pensamos passa a ser aquilo que *nós dizemos que estamos pensando* – e isto pode não corresponder aos nossos comportamentos. Essa mudança radical – que teria ocorrido com nossos ancestrais – seria não só o surgimento da crença de que temos um acesso privilegiado aos nossos pensamentos como também a crença de que somos a autoridade máxima acerca de nossos pensamentos, mesmo que esses estejam em franca contradição com o comportamento que poderia corresponder a eles. Isto significa dizer que nossos relatos introspectivos são sempre necessariamente corretos – mesmo quando houver evidências (comportamentos) contrárias a eles.

O passo seguinte na alegoria de Jones é estender o acesso privilegiado e a primazia do relato introspectivo à percepção. Aplicando a teoria de Jones, nossos ancestrais mudaram a natureza de seus episódios de percepção. Antes dele, nossos ancestrais concebiam o episódio perceptual de, por exemplo, ver que um triângulo é vermelho como intrinsecamente ligado a um comportamento verbal relatando que o triângulo era vermelho. Falar de um triângulo vermelho (comportar-se verbalmente em relação a ele) era praticamente a mesma coisa que ver um triângulo vermelho. Linguagem e percepção caminhavam sempre juntas; perceber um objeto e falar dele eram ações praticamente inseparáveis. Introduzindo a ideia de pensamento e de entidades não observáveis como subjacentes a suas próprias impressões, eles deixam de ver diretamente o triângulo vermelho e passam *a relatar a experiência* de que estão vendo um triângulo vermelho. Ver um triângulo vermelho passa a ser uma experiência interna, mediada pelo pensamento e não uma experiência sensorial imediata ligada a um tipo de comportamento – seja ele verbal ou não. Jones introduz então mais uma entidade inobservável na sua teoria, as *impressões*, que seriam um tipo especial de pensamento que expressa episódios perceptuais. As impressões são correlatos mentais de estados perceptuais de uma pessoa – estados mentais aos quais essa pessoa tem acesso privilegiado. Aos poucos, os correlatos mentais passam a prevalecer sobre os próprios episódios perceptuais da pessoa e suas consequências em termos de comportamentos. Um triângulo vermelho passa a ser visto como tal somente na medida em que os integrantes dessa comunidade passam a partilhar esses estados perceptuais e não apenas exibir os comportamentos e sentenças que antes (quando linguagem e ação não estavam separadas) eram proferidas na presença de um triângulo vermelho.

O cenário está, então, completo. Jones – que não é apenas um filósofo, mas também um pregador, o *reverendo* Jones – passa a espalhar a boa-nova e treinar seus conterrâneos para assimilar sua teoria. Esse treinamento consiste basicamente em convencer seus companheiros da primazia dos relatos introspectivos sobre as evidências comportamentais. Para isto, duas técnicas foram adotadas. Em primeiro lugar Jones procurou convencer seus conterrâneos da inexistência de qualquer ligação conceitual ou lógica entre esses estados internos inobserváveis (os pensamentos) e seus comportamentos correspondentes. Com isto, abria-se o espaço necessário para considerar esses pensamentos como dotados de uma existência independente, autônoma. A segun-

da técnica consistiu em convencer seus companheiros – em caráter definitivo – da verdade primeira, ou seja, da primazia dos relatos introspectivos sobre qualquer evidência comportamental contrária. Feito isto, essas entidades inescrutáveis (os pensamentos) passaram a ganhar vida própria, ou seja, uma "realidade efetiva", que não se resumiria a uma simples suposição teórica ou uma hipótese. A teoria se sobrepôs à realidade e deixou de ser teoria, ou seja, passou a ser mais real que o comportamento e a percepção. Ao final do treinamento, se houvesse uma evidência comportamental que conflitasse com o relato introspectivo de um dos membros da comunidade, prevaleceria este último. Assim sendo, quando os membros dessa comunidade passaram a acreditar que o que eles estavam pensando era aquilo que eles *supunham* ou *diziam* estar pensando e o relato introspectivo passou a ser tomado como autoevidente, surgiu a ideia de *mente*.

A ideia de mente surgiu de uma inversão fundamental propiciada pela expansão da linguagem – e essa teria sido a grande tarefa realizada pelo reverendo Jones. Segundo a alegoria de Sellars, a noção de mente foi engendrada pela expansão da linguagem que propiciou o triunfo dos relatos introspectivos sobre o comportamento e sobre a percepção. Palavras e relatos introspectivos tornaram-se, ao longo desse treinamento, mais reais do que o mundo observável. A própria ideia de "primeira pessoa" e de "acesso privilegiado" teriam sido forjadas pela linguagem. O treinamento ao qual Jones teria submetido sua comunidade teria sido transmitido a seus descendentes até chegar a nós, formando comunidades que acreditam que mentes não seriam apenas uma invenção linguística. A teoria de Jones teria se consolidado na forma de psicologia, uma disciplina que, entretanto, herdaria todas as dificuldades de tratar a mente como sendo uma realidade e não apenas um incidente produzido por uma expansão exagerada da linguagem. Essas dificuldades se expressariam, inevitavelmente, na incapacidade da psicologia – e mais tarde da própria filosofia da mente – em relacionar mentes com comportamentos.

Mas a alegoria de Sellars não termina aqui. Jones teria sido bem-sucedido num primeiro momento ao fazer prevalecer o relato introspectivo sobre o comportamento e inventar a mente. Só por isto ele teria razões de sobra para se orgulhar de sua façanha. Contudo, num estágio posterior ele teria querido refinar sua teoria e transformá-la numa autêntica ciência. Ele teria querido encontrar correlatos objetivos, neuronais, de suas entidades inobserváveis e tornar sua teoria uma ciência do cérebro. Mas, ao voltar-se para essa nova tarefa, ele já não podia mais desvencilhar-se da própria linguagem que ele criara – a linguagem com todas as expansões para uma semântica e para uma teoria do comportamento baseada em entidades inobserváveis. Essa linguagem o afastara tanto do mundo e da percepção direta desse que já não era mais possível confiar nela. Nem para dizer o que de fato existe, nem para dizer como o mundo é. Talvez não fosse mais possível fazer nenhum tipo de ciência a partir dessa linguagem, mas agora era tarde demais. Jones se achou num caminho sem volta – um caminho que fez com que o mito da mente prevalecesse até hoje.

Muitas pessoas poderiam achar chocante esse mito relatado por Sellars, outras poderiam achá-lo até trivial. Na verdade, tentativas de desfazer a noção de mente e con-

cebê-la como o resultado de alguma armadilha armada por nossa própria linguagem – uma armadilha que nos confinaria a um problema insolúvel – já existiam na filosofia da mente. Muito antes de Sellars, na década de 1940, essa ideia foi sugerida pelo filósofo inglês Gilbert Ryle (1900-1976) cujas principais concepções analisaremos a seguir.

Gilbert Ryle[1]

Imagine que um dia alguém lhe peça para conhecer a Universidade de São Paulo. Gentilmente, você coloca essa pessoa no seu carro e começa a dar uma longa volta, parando na frente de cada edifício e dizendo: "Esse é o prédio da medicina", "esse é o prédio da psicologia" e assim por diante. No final do dia você teria mostrado a essa pessoa todos os prédios do campus, um por um. Imagine agora se essa pessoa, ao descer do carro e se despedir de você, agradecendo a gentileza, dissesse: "Muito obrigado, você me mostrou tantos prédios! Mas você ainda não me mostrou a Universidade de São Paulo".

Certamente esta última frase causaria perplexidade. O que essa pessoa está querendo que eu mostre? Por acaso você já não teria mostrado para ela a Universidade de São Paulo? Afinal, o que será que ela está querendo *ver*? Será que ela pensa que a Universidade de São Paulo deve ser um outro prédio para além dos prédios de todas as escolas que eu já mostrei? Ou será que ela pensa que, para além de todos esses outros prédios, deve existir *alguma outra coisa* que caracteriza a Universidade de São Paulo?

Imagine agora que o convidem para visitar um casal amigo seu. Você sabe que eles são casados. Você chega na casa deles e começa a notar a casa, os quartos, os filhos, os brinquedos dos filhos, dois carros na garagem e assim por diante. Contudo, há uma coisa que lhe causa frustração: você não consegue enxergar o *casamento* deles. Onde estaria o casamento deles? Seria um pedaço de papel passado em cartório e guardado em algum arquivo? Também não. Esse pedaço de papel *representa* e atesta o casamento deles, mas não *é* o casamento deles.

A primeira situação ilustra o que constitui, para Ryle, um dos principais equívocos da filosofia da mente. Inspirados pela tradição cartesiana, os filósofos da mente começaram a supor que, para além de todo um conjunto de comportamentos e disposições que observamos nas pessoas, existiria algo mais, algo como uma substância subjacente a todas essas manifestações, da mesma maneira que a pessoa que você levou para conhecer a Universidade de São Paulo perguntou: "mas afinal, onde está a Universidade de São Paulo?" Da mesma maneira que a Universidade de São Paulo se esgota no conjunto de prédios que a compõem, a mente se esgota no conjunto de comportamentos e disposições manifestados pelas pessoas.

1. Algumas passagens desta seção apareceram em Teixeira (1996a). Os exemplos são de Dennett (1969).

Supor que existe algo mais do que isto é o equívoco que Ryle aponta. Um equívoco que leva a supor que, para além das partes deve haver algo subjacente a elas – algo que os cartesianos identificariam como sendo uma substância. Uma substância com propriedades especiais, dentre elas, a imaterialidade que a tornaria incompatível com a produção causal dos comportamentos. Supor a existência dessa substância é incorrer numa ilusão; é postular a existência de um fantasma na máquina (*the ghost in the machine*). Um fantasma que, uma vez postulado, leva a todas as dificuldades e problemas insolúveis acerca da natureza da mente e sua relação com o cérebro e com o mundo físico. É o mesmo que postular a existência da quintessência cartesiana de que falávamos no capítulo II: a quintessência que necessariamente estaria faltando aos autômatos e que impediria, em princípio, que eles pudessem replicar uma vida mental igual à nossa. É supor que vida mental e consciência *são algo mais*, uma essência ou *élan* intangível para além da possível replicação dos comportamentos e disposições que a caracterizam.

A segunda situação ilustra o que significa confundir um conceito com uma coisa. Ou confundir um conceito com algum tipo de substância que deveria existir em algum lugar. Casamento é um conceito que designa um tipo de relação entre duas pessoas; uma relação que não é nem palpável nem observável. Podemos apenas observar suas consequências ou suas manifestações. É isto que designamos quando falamos de casamento. Ryle diria que a mente não é nada além de um conceito: um conceito que utilizamos para designar um conjunto de comportamentos e disposições exibidos pelas pessoas. E também para designar um determinado tipo de organização que inferimos a partir desses comportamentos e disposições. Mas mente não é uma coisa, nenhuma substância física. Tampouco seria uma substância imaterial que, como um fantasma dentro da máquina, seria responsável por essa organização.

Supor que da observação de um conjunto de comportamentos organizados podemos inferir a existência de algum tipo de *entidade subjacente* é incorrer no erro cartesiano. Um erro tornado ainda maior, pois pressupõe que a um conceito deva necessariamente corresponder a existência de algum tipo de entidade. Alguns tipos de conceitos – como seria o caso do conceito de mente – designam relações e não entidades. Relações que não levam a nada tangível, como, por exemplo, a ideia de "quarta-feira". "Quarta-feira" não designa nada tangível, apenas uma relação que estipulamos entre os diversos dias da semana e que podemos apenas convencionalmente representar num calendário. Saltar de um conceito relacional ou de algo que designa uma organização para uma entidade é uma extravagância ontológica, ou seja, é povoar nossa filosofia com excrescências das quais se originam os mais indesejáveis pseudoproblemas.

A tarefa da filosofia da mente será então estirpar as extravagâncias e com elas dissolver os pseudoproblemas. É a tarefa de exorcizar o fantasma da máquina, mostrando que ele é apenas uma ilusão. Para isto precisamos saber a origem dessas extravagâncias ou fantasmas. E Ryle supõe que a origem de todas essas confusões está na linguagem e no modo como a empregamos, criando um vocabulário psicológico que nos induz a

supor a existência dessa substância imaterial. Quando isto ocorre, temos aquilo que Ryle batizou de "erro categorial" (*category mistake*) – um tipo de erro que requer uma clarificação ou uma terapia da linguagem.

A estratégia dessa terapia consiste em separar nitidamente o vocabulário físico do vocabulário mental. Essa triagem seria o passo inicial para verificar como se originou o fantasma na máquina para, em seguida, livrar-nos dele. Ou seja, ao usar inadvertidamente nossa linguagem cotidiana, frequentemente transpomos termos de um vocabulário físico e os aplicamos na construção de um vocabulário mental, gerando, com isso a ilusão implícita de que o mental é uma entidade ou algum tipo de substância com existência independente.

No seu livro publicado em 1949, *The Concept of Mind* [O conceito de mente], Ryle dá vários exemplos de como usar essa estratégia de estirpar o fantasma da máquina pela eliminação das transgressões categoriais. Por exemplo, a sentença " hoje estou muito cansado mentalmente" ou "minha mente está cansada hoje" seriam exemplos típicos de transgressões categoriais. Mentes não ficam cansadas, são os nossos organismos que ficam cansados após a realização de atividades intelectuais que requerem uma vigília prolongada. Músculos se cansam após muita atividade, mas nem mentes nem cérebros têm músculos. Quando dizemos "minha mente está cansada hoje" estamos usando uma metáfora que, se tomada literalmente, gera uma transgressão categorial. Da mesma maneira, falamos, por exemplo, que "estamos ouvindo uma música na nossa mente", ou "sua voz ficou marcada na minha mente" como se a ideia de "na minha mente" designasse um lugar físico ou como se mentes tivessem algum tipo de localização física, um palco real que poderia ser ocupado por sons e imagens. Essas metáforas, quando se tornam literais (talvez pelo uso constante) geram a ideia de que a mente deve ser algo físico ou algum tipo de substância subjacente a meus comportamentos e disposições para agir dessa ou daquela maneira.

Ora, esse programa de terapia linguística proposto por Ryle parece bastante tentador. Mas será exequível? A questão que se coloca é: será sempre possível fazer essa separação ou haverá alguns termos híbridos, inerentes ao modo como empregamos a nossa linguagem cotidiana que resistiriam a esse tipo de terapia? Por exemplo, quando falamos do Estado de São Paulo ou dos gols marcados num jogo de futebol estaremos falando de algo físico ou de algo mental?[2] Haveria outros casos que também poderíamos classificar como anômalos, ou seja, expressões linguísticas situadas de forma híbrida entre o físico e o mental?

Tomemos por exemplo as noções de distância e de *medida* de distâncias, tais como quilômetros e metros. Será possível situar a ideia de quilômetro no vocabulário físico? Ou devemos situá-la no vocabulário mental? Podem os quilômetros que existem entre a Terra e a Lua serem identificados com algo *físico* do mundo? Certamente que não.

2. Esses dois exemplos foram adaptados de Searle (1992).

Mas, por outro lado, não é possível conceber a distância a não ser como algo físico. Quilômetros e outras medidas, como, por exemplo, graus centígrados, teriam uma existência tênue entre o físico e o mental – uma existência mais tênue do que aquela dos pensamentos, crenças e desejos. Certamente não podemos borrifar tinta num pensamento, numa crença ou num desejo, mas podemos dizer que um pensamento ocorre na minha cabeça – e ocorrer no espaço ou em algum lugar certamente constitui uma propriedade física. Uma dor pode ser intensa, da mesma forma que afirmamos que uma chama de fogão é intensa. Um desejo pode embrulhar meu estômago, o mesmo que ocorre quando como um sanduíche de alho.

Entretanto, a análise linguística dificilmente poderia classificar como transgressões categoriais as afirmações de que o pensamento ocorre na minha cabeça, que a dor é intensa ou que um desejo embrulha meu estômago: essas sentenças não só fazem sentido como o fazem precisamente por transitar entre o físico e o mental. O resultado é inverso no caso dos quilômetros e outras distâncias ou medidas, nos quais a análise linguística reverte nossa tendência habitual de situá-los do lado do vocabulário físico.

Um outro exemplo interessante é a análise do termo "voz". Devemos situar vozes no vocabulário físico ou no vocabulário mental? Quando afirmo "ouço uma voz" ou "perdi minha voz" ou mesmo "ele ouve vozes" será que posso situar "voz" em domínios distintos? Que sentido tem tratar "voz" como coisa física quando afirmo "perdi minha voz"? Terá sentido a afirmação "perder a voz" entendida como perder um *objeto físico*? Por outro lado, a ideia de voz como coisa física não pode ser abandonada quando gravo minha voz numa fita magnética e a mando para um amigo em Londres. Será então "a voz que perco" uma metáfora que significa, na realidade, a perda temporária de uma disposição? Nesse caso, para resolver essa dificuldade teríamos apenas de fazer um levantamento dos diferentes sentidos da palavra "voz": haveria uma alternância entre o sentido físico e o sentido mental. Mas isto rapidamente leva a paradoxos, pois, se assim fosse, a voz que eu emito não poderia ser a voz que eu perdi ontem; uma seria física e outra seria mental. Talvez o mesmo se aplicasse às vozes que ouço quando tenho um surto psicótico: não *penso* vozes, *ouço* vozes, e como poderia saber se elas são físicas ou mentais?

Esses casos-limite nos colocam numa espécie de dilema. Ou abandonamos o programa ryleano de separar o vocabulário físico do vocabulário mental e com ele seu pressuposto fundamental de que é sempre possível detectar e eliminar as transgressões categoriais ou admitimos que a passagem do físico para o mental, expresso em sentenças híbridas, não leva necessariamente a um uso indevido da linguagem e à geração de paradoxos.

O materialismo eliminativo dos Churchlands[3]

Se os trabalhos de Sellars e de Ryle têm como proposta desfazer mitos e exorcizar fantasmas – mostrando que a ideia de mente nada mais seria do que uma extravagância da linguagem –, o dos eliminativistas segue uma inspiração mais radical. Trata-se não apenas de identificar os subúrbios da linguagem que confeririam ao mental um estatuto ontológico que esse não possui. Dá-se um passo mais do que a identificação crítica proporcionada pela terapia linguística: é preciso decretar, desde o início, a falência dessa ontologia pelo reconhecimento da inadequação do vocabulário psicológico cotidiano para descrever o mental e substituir a imagem comum da mente por uma imagem científica derivada da neurociência. O vocabulário psicológico cotidiano seria incompatível com o discurso da ciência e, por isso, sua permanência seria, igualmente, intolerável no interior de uma visão científica do mundo. Isto marca uma diferença de estratégias e de objetivos entre, por um lado, Sellars e, por outro, os materialistas eliminativos, embora seus projetos possam, de modo geral, ser inscritos no horizonte comum de desfazer a ideia de mente.

O materialismo eliminativo pode ser considerado uma radicalização do projeto reducionista, no intuito de superar algumas das dificuldades enfrentadas pelas tentativas de redução. Embora tenha sido proposto já na década de 1960 por Paul Feyerabend e também por Richard Rorty, foi sobretudo a partir da década de 1980 – com o famoso casal Churchland (Paul e Patricia Churchland) – que ganhou força, gerando calorosas discussões na filosofia da mente. Com vistas a uma caracterização geral do movimento, vamos nos concentrar, portanto, na proposta dos Churchlands.

Talvez a melhor maneira de compreendermos a proposta eliminativista seria ressaltar a diferença entre redução e eliminação. Tradicionalmente, quando utilizamos o termo "redução" na literatura científica e filosófica, estamos caracterizando sobretudo uma relação entre teorias, onde uma velha teoria T1 é reduzida logicamente a uma nova teoria T2 e os eventos antes explicados por T1 passam a ser explicados por T2. Assim, temos um caso exemplar na história da física, em que a temperatura, antes explicada pelas leis da termodinâmica clássica, passou a ser entendida em termos de energia cinética molecular, o que garantiu a redução da termodinâmica clássica à mecânica estatística.

Os objetivos últimos do ideal reducionista são a unificação explicativa e a simplificação ontológica, embora essa última nem sempre seja pretendida. Entretanto, no caso dos fenômenos mentais, encontramos frequentemente a tentativa de efetuar essa redução ontológica, na afirmação de que eles são idênticos a eventos cerebrais. Daí a busca de correlatos neurais para todo estado mental e a esperança de que no futuro a neurociência nos proporcionará uma taxonomia que garanta uma correspondência estrita com a taxonomia de nosso senso comum, para que a redução seja bem-sucedida.

---

3. Na elaboração desta seção, bem como da seguinte, contei com a colaboração direta do Prof. Saulo de Freitas Araújo.

Tendo em vista as enormes dificuldades encontradas na realização desse ousado empreendimento, o ponto de partida dos Churchlands é a recusa daquilo que eles consideram um erro fundamental do projeto reducionista tradicional: a suposição de que nossa linguagem psicológica utilizada habitualmente para explicar e predizer o comportamento humano (*folk psychology*) é adequada. A *folk psychology* ou "psicologia popular" seria uma espécie de teoria habitual que todos nós possuímos, através da qual explicamos os comportamentos de outros seres humanos recorrendo às ideias comuns de "intenção", "crença", "desejo" e outros termos que compõem o chamado vocabulário mentalista. Segundo os Churchlands, nós não precisamos buscar uma redução dessa teoria inadequada – a *folk psychology* – a uma eventual neurociência amadurecida, mas simplesmente uma eliminação da primeira, dado que ela é falsa. No entanto, é importante ressaltar que não se trata aqui de uma eliminação do mental, mas tão somente de uma linguagem mentalista, uma vez que os Churchlands não negam a realidade de nossa experiência subjetiva.

A proposta de uma reforma da linguagem da psicologia adequando-a ao avanço das teorias neurobiológicas seria uma consequência natural da eliminação progressiva do vocabulário mentalista da *folk psychology*. Esse projeto percorre todas as versões do materialismo eliminativo, estando presente inclusive naquelas que antecederam o trabalho dos Churchlands. Rorty (1965) sustentou que toda linguagem é um produto de interações sociais e que, portanto, não existem termos naturalmente dados. Sendo assim, não haveria razão *a priori* para excluir a possibilidade de que no futuro nós venhamos a nos referir a estados cerebrais em vez de utilizarmos os termos mentalistas que hoje impregnam nossa linguagem. Todo o vocabulário mentalista que empregamos hoje seria fruto de um longo aprendizado, transmitido durante várias gerações pelos nossos ancestrais. Assim, poderíamos perfeitamente ser treinados para falar uma outra linguagem, na qual os termos básicos fossem estados cerebrais, que seriam, ao mesmo tempo, públicos e privados. Posteriormente, Rorty abandonou essa posição e, com ela, sua posição favorável ao materialismo eliminativo. O mesmo não ocorreu, entretanto, com os Churchlands, que continuaram defendendo o materialismo eliminativo e a proposta de uma reforma linguística através da eliminação da *folk psychology* e dos termos mentalistas tradicionais. O "neurologuês" tornar-se-ia, num futuro não muito distante, a genuína linguagem da psicologia.

Outra característica fundamental do materialismo eliminativo dos Churchlands, que tem gerado alguns erros de interpretação de sua proposta, é que eles não recusam a possibilidade de uma futura teoria psicológica ser desenvolvida juntamente com uma teoria neurobiológica, até que uma redução da primeira em relação à segunda se torne possível. Isso quer dizer que eles aceitam a redução interteórica, desde que a teoria psicológica seja diferente da *folk psychology*. Em outras palavras, os eliminativistas mantêm a perspectiva reducionista, adotando o eliminativismo apenas nos casos em que a teoria for inadequada.

Tendo evidenciado, então, o alvo dos ataques dos Churchlands, podemos formular a seguinte pergunta: mas o que há de errado com a *folk psychology*? Para desacredi-

tar nossa linguagem psicológica de senso comum e condená-la ao desaparecimento, os defensores do materialismo eliminativo também recorrem à história da ciência, como no caso da redução interteórica, mostrando casos de eliminação ontológica de velhas teorias em favor da ontologia de uma nova e superior teoria. Nesse sentido, um dos exemplos mais citados é a teoria do flogisto, utilizada para explicar fenômenos como a combustão e a ferrugem. Acreditava-se que quando um pedaço de madeira queima ou uma barra de metal enferruja, isso acontece pelo fato de haver a liberação de uma substância inerente aos corpos chamada flogisto. Mais tarde descobriu-se que ambos os processos ocorrem não devido à perda de alguma coisa, mas sim porque os corpos ganham uma substância advinda da atmosfera, a saber, o oxigênio. Dessa forma o conceito de "flogisto" não foi identificado ou reduzido a nenhum outro conceito da nova teoria do oxigênio, mas foi eliminado da ciência, em função de se referir a algo que não existe.

Um outro exemplo mencionado pelos eliminativistas, já mais próximo à psicologia, é o da possessão demoníaca. Em séculos passados, casos de psicose eram considerados uma manifestação do espírito do demônio, que se incorporava nas pessoas. Da mesma forma, considerava-se seriamente a existência de bruxas por toda a parte, responsáveis por comportamentos socialmente indesejáveis. Entretanto, com o avanço de pesquisas e de novas teorias sobre a disfunção mental, ambas as entidades foram eliminadas da ontologia científica, devido à sua inadequação teórica.

Com base nesses paralelos históricos, os defensores do materialismo eliminativo afirmam que o destino dos conceitos pertencentes à *folk psychology* – desejo, crença, intenção, medo, esperança, sensação etc., – será rigorosamente o mesmo, devido à sua estagnação e também à sua incapacidade de explicar vários fenômenos da vida mental, como, por exemplo, o sono, as doenças mentais, a aprendizagem etc. Tão logo a neurociência se desenvolva e alcance um alto grau de maturidade, a inadequação de nossas concepções atuais tornar-se-á visível e seremos então capazes de desenvolver um modelo conceitual compatível com o conhecimento neurocientífico, que nos permita explicar verdadeiramente nossas atividades mentais.

Para levar adiante seu projeto eliminativista, no entanto, os Churchlands tinham que superar um obstáculo fundamental: a herança cartesiana. Como foi ressaltado no capítulo II, o legado que a tradição inaugurada por Descartes nos deixou gira em torno de dois temas principais: a intencionalidade como marca distintiva do mental – que gera o problema do conteúdo ou do significado – e a intransponibilidade da perspectiva de primeira pessoa, exemplificada pelo caso dos *qualia* (qualidades da experiência subjetiva), que envolve a consciência.

Retomando a caracterização da intencionalidade feita anteriormente, dizemos que um estado mental é intencional porque se refere a algo, podendo esse algo existir ou não no mundo. Por exemplo, quando penso numa "mula sem cabeça", o conteúdo do meu pensamento é uma "mula sem cabeça", embora "mulas sem cabeça" não existam. Essa capacidade de se referir a um conteúdo qualquer seria, então, a característica distintiva e irredutível dos fenômenos mentais, exibida pelos assim chamados *estados*

intencionais ou atitudes proposicionais (crenças, desejos, pensamentos etc.). Nesses estados, o significado ou conteúdo é expresso por uma proposição P específica (x acredita que P, x deseja que P, x pensa que P etc.).

O argumento central dos eliminativistas é que a intencionalidade de modo algum constitui uma refutação do materialismo, uma vez que estados puramente físicos, como os estados cerebrais, também possuem conteúdo proposicional, sendo, portanto, intencionais. Para exemplificar seu ponto de vista, os Churchlands apresentam uma comparação entre certas atitudes proposicionais da *folk psychology* e o que eles chamam de atitudes numéricas da física. Assim, da mesma forma que "x acredita que P", "y tem uma extensão de *n* metros", bem como "x deseja que P" e "y tem uma velocidade de *n* metros por segundo" ou, ainda, "x pensa que P" e "y tem uma energia cinética de *n* jaules". A única diferença entre esses dois tipos de "atitude" é que, no primeiro caso, a variável deve ser substituída por uma proposição específica e, no segundo, por um número, a fim de que tenhamos um verdadeiro predicado. Em ambos os casos, porém, haveria intencionalidade.

Não podemos deixar de ressaltar que é a teoria semântica dos Churchlands que lhes permite neutralizar os argumentos antimaterialistas. Para os eliminativistas, ter um conteúdo ou significado é apenas uma questão de desempenhar um papel específico numa complexa rede inferencial ou computacional, que nos permite identificar certas relações abstratas (relações lógicas e numéricas, por exemplo) entre as diferentes atitudes e postular leis gerais. Por exemplo, a partir das noções de implicação e de consistência lógicas, eu posso postular a seguinte relação entre duas atitudes preposicionais, em forma de lei: se x teme que P, então x deseja que não P. Desse modo, se os estados cerebrais ou computacionais podem desempenhar um papel numa dinâmica inferencial, exibir relações abstratas e gerar leis, então eles possuem conteúdo e, consequentemente, intencionalidade.

O segundo grande obstáculo encontrado pelos eliminativistas é conhecido na literatura como o problema dos *qualia*, isto é, os aspectos qualitativos de nossas experiências subjetivas. Nesse sentido, um dos argumentos mais tradicionais foi apresentado pelo filósofo Thomas Nagel, em seu clássico artigo *What is it like to be a bat?* [O que é ser como um morcego?], que examinamos no capítulo anterior. Contra os Churchlands, Nagel afirma que os *qualia* de nossas sensações nos são revelados apenas através da introspecção, diferentemente das informações obtidas pela neurociência. Além disso, nós poderíamos conhecer tudo sobre a neurofisiologia do morcego e sua interação com o mundo físico, mas nunca saberíamos e nem seríamos capazes de imaginar como é ser um morcego. Baseado nesse argumento, Nagel sustenta que as pretensões reducionistas estariam condenadas ao fracasso, uma vez que os *qualia* ficariam ausentes da abordagem objetiva da neurociência.

Outro argumento muito semelhante – que também examinamos no capítulo anterior – é o de Frank Jackson, que apareceu no seu artigo de 1982, chamado *Epiphenomenal qualia* [*Qualia* epifenomenais]. Jackson conta a história de uma brilhante neurocientista chamada Mary, que viveu toda a sua vida presa dentro de um quarto, onde

recebia informações sobre o mundo exterior através de livros e de um monitor preto e branco. Sua experiência visual se resumia a tonalidades de preto, branco e cinza. No entanto, Mary sabia tudo sobre as estruturas físicas do cérebro e de seu sistema visual, assim como de seus estados reais e possíveis. De acordo com Jackson, haveria porém uma coisa que ela não sabia e não podia imaginar sobre as experiências subjetivas de outras pessoas que viviam fora de seu quarto e sobre suas possíveis experiências quando fosse deixá-lo: a natureza da sensação de ver um tomate bem vermelho. Dessa história Jackson conclui que mesmo um conhecimento completo dos aspectos físicos da percepção visual e da atividade cerebral a ele relacionada seria insuficiente para explicar todos os fenômenos mentais. Portanto, o materialismo reducionista seria falso.

Segundo Paul Churchland, um dos grandes problemas do argumento de Nagel é que ele comete uma petição de princípio. Quando afirmo que meus *qualia* são conhecidos por mim através da introspecção e que meus estados cerebrais não o são, eu já estou pressupondo aquilo que o argumento deveria provar, a saber, que estados mentais não são idênticos a estados cerebrais. Ora, se os estados mentais forem de fato idênticos a estados cerebrais, então o conhecimento de meus *qualia* é o conhecimento dos estados cerebrais a eles correspondentes. O que pode variar, diz Churchland, é apenas o modo de descrição. Eu posso ou não descrever meus estados mentais em termos cerebrais, dependendo das informações de que eu dispuser. Em outras palavras, a identidade pode ser um aspecto do mundo real, independentemente de eu conhecê-la.

Em relação à "subjetividade" do morcego, afirma Churchland, é de fato possível que nós não tenhamos acesso a alguns de seus estados internos. Mas isso não implica, de modo algum, que os *qualia* do morcego não sejam estados físicos e que, portanto, o fisicalismo seja falso. Trata-se apenas de uma limitação da capacidade humana, derivada das diferenças estruturais específicas entre o nosso cérebro e o do morcego.

Por fim, no que diz respeito ao caso de Mary, Churchland aponta um passo injustificado no raciocínio de Jackson, quando ele supõe que Mary, mesmo sabendo tudo o que há para saber sobre o cérebro, não saberá como é ter uma experiência do vermelho. Nós não temos ideia do que um conhecimento neurocientífico total poderia nos revelar. É bem provável que ele provocaria uma mudança radical na nossa maneira de conceber os fenômenos mentais. De qualquer forma, conclui Churchland, trata-se de uma questão empírica e que, portanto, não pode ser defendida *a priori*, como no argumento de Jackson.

O futuro do materialismo eliminativo

O primeiro ponto importante a se considerar na avaliação das propostas eliminativistas é o seu caráter programático. Trata-se de uma aposta no desenvolvimento futuro da neurociência e na sua capacidade de nos fornecer uma explicação mais adequada dos fenômenos mentais. Entretanto, estamos ainda muito longe de uma teoria neurobiológica e de uma teoria psicológica abrangentes, para que a redução seja efetuada e a

folk psychology eliminada. Portanto, a fim de não cometermos injustiças na avaliação, é preciso fazer uma distinção fundamental entre o materialismo eliminativo e a neurociência: trata-se, no primeiro caso, de uma teoria filosófica da mente e, no segundo, de uma ciência do cérebro. Existe uma diferença quanto à natureza das investigações, muito embora os desenvolvimentos futuros possam convergir.

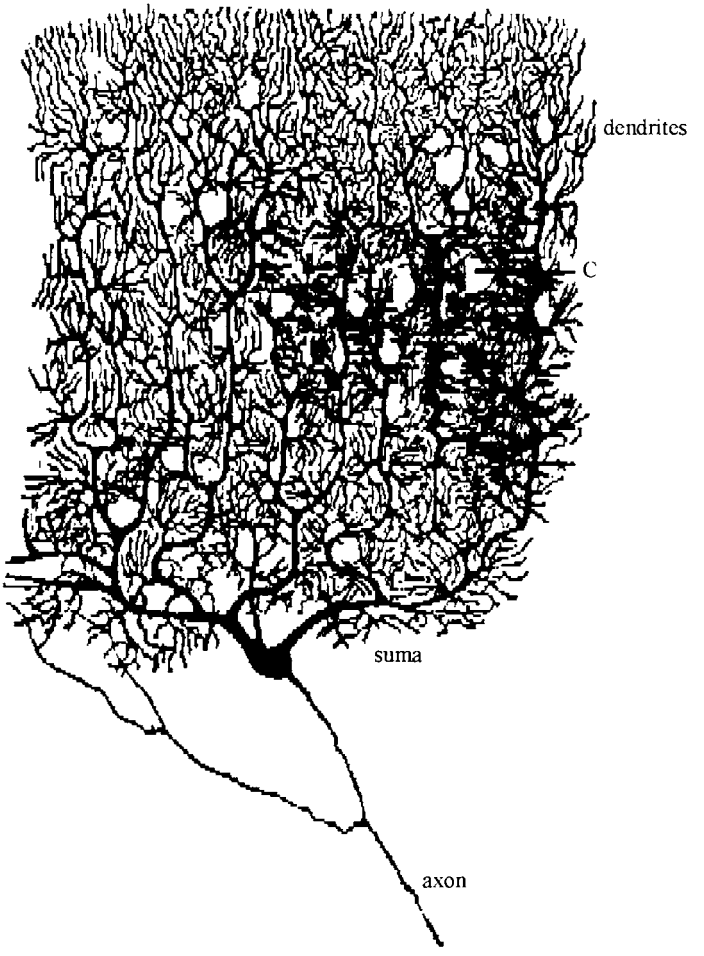

Fig. 5.1 – Célula de Purkinje, numa ilustração de Ramón y Cajal. Note-se a extraordinária complexidade das ramificações.

Há uma questão, contudo, que devemos ter em mente desde o princípio, independentemente de qualquer progresso empírico: pode a neurociência solucionar os principais problemas colocados pela filosofia da mente? Essa questão ganha relevância na medida em que o sucesso do materialismo eliminativo depende da possibilidade de alcançarmos um conhecimento completo do cérebro. Em outras palavras, a tão sonhada eliminação da *folk psychology* só se supõe possível a partir de um mapeamento integral entre o

mental e o cerebral. Caso contrário, todo o programa eliminativista estará ameaçado, uma vez que restarão algumas categorias da *folk psychology* sem correspondência cerebral específica. Assim, o sonho de uma linguagem puramente cerebral seria frustrado. De fato, as grandes dificuldades do programa eliminativista parecem se derivar exatamente dessa dependência em relação a uma neurociência completa.

Ao tentar superar esse obstáculo, o eliminativista começa a se deparar com uma série de problemas, começando pela imensa complexidade do cérebro. Estima-se que no nosso cérebro haja cerca de 10^{13} neurônios e 10^{15} sinapses, com alguns desses neurônios como, por exemplo, as células de Purkinje – podendo ter ramificações muito complexas.

A complexidade do cérebro humano pode realmente tornar-se um grande obstáculo à realização do mapeamento total. Se o número provável de neurônios (n) for mesmo 10^{13} e entre eles houver combinações binárias (do tipo "sim" ou "não"), o número de conexões entre eles subirá para n^2 – o que já é um número alarmante. Mas é bem provável que as conexões entre os neurônios não sejam apenas do tipo binário, o que tornaria esse número n ainda mais elevado. Sendo assim, como poderíamos reproduzir esse número astronômico de neurônios e sinapses, a fim de não comprometer o programa eliminativista?

Uma saída aparentemente sensata para esse problema seria a utilização de um supercomputador, onde instalaríamos um poderoso software gráfico que "desenharia" os 10^{13} neurônios para, em seguida, traçar as 10^{15} sinapses e representar as n^2 configurações que poderiam ocorrer nesse modelo, caso as conexões sejam apenas binárias. Ora, quantas operações esse computador teria que realizar até que se chegasse a algo parecido com um modelo do cérebro? No mínimo 10^{28} operações – um número que é igual ao número de partículas físicas que se estimam existir no universo. Mesmo um supercomputador extremamente veloz, funcionando a uma velocidade próxima à da luz (que é a velocidade-limite do universo) levaria alguns milênios para completar sua tarefa. Em outras palavras, estamos diante de um sério problema de *complexidade computacional.* Curiosamente, o próprio Paul Churchland sugere – no seu livro *The Engine of Reason, the Seat of the Soul* [O mecanismo da razão e o sítio da alma] publicado em 1995 – que o número de conexões possíveis entre os neurônios do cérebro humano iguala-se ao número de partículas no universo.

O problema pode se agravar ainda mais. Estivemos até agora supondo que a distribuição dos neurônios e das conexões entre eles estariam obedecendo a algum tipo de regra – uma regra constante a partir da qual podemos construir esse software gráfico. No entanto, não sabemos se a distribuição dos neurônios e de suas conexões obedece a algum tipo de regra que pudesse ser reproduzida por um software. E caso não haja essa regra (ou esse conjunto de regras), a construção de um modelo de cérebro através de um software tornar-se-ia inviável, ou seja, estaríamos diante de um problema não computável. Portanto, o apelo dos eliminativistas a uma neurociência computacional seria inútil[4].

---

4. Para outras apresentações desse argumento cf. Teixeira (1998a, 1998b).

Até aqui, estivemos falando apenas de problemas de natureza técnica. Mas o eliminativista enfrenta ainda dificuldades filosóficas. Em primeiro lugar, poderíamos aplicar ao materialismo eliminativo o argumento que Popper apresenta em seu livro *The Self and its Brain* [O eu e seu cérebro] (1977), escrito em parceria com o neurofisiólogo John Eccles. O que Popper quer demonstrar é que qualquer tipo de materialismo radical se autoanula, em função de não poder sustentar sua validade com argumentos racionais. Ora, se todo estado mental é idêntico a um estado cerebral, então a proposição "o materialismo é uma teoria verdadeira" é apenas um produto físico-químico do cérebro, assim como a proposição "o materialismo não é uma teoria verdadeira". Sendo assim, afirma Popper, todos os padrões de validade tornam-se contingentes e, portanto, não servem para discriminar teorias contrárias[5].

Uma outra dificuldade filosófica que o materialismo eliminativo enfrenta está diretamente relacionada ao problema da intransponibilidade da perspectiva de primeira pessoa. Consideremos – como já o fizemos no capítulo I – um grupo de 10 pessoas submetidas a sessões de neuroimagem para a identificação de padrões de atividade cerebral relacionada ao pensamento. Ao final das sessões, podem essas pessoas reconhecer seus próprios cérebros, somente a partir da apresentação das imagens obtidas? O que essas imagens indicam a respeito dos conteúdos dos pensamentos de cada um? A lição a ser extraída desse exemplo é que a neurociência não pode ignorar a perspectiva subjetiva, se quiser explicar nossa vida mental. No entanto, isso parece nos levar a um grande paradoxo, a saber, como pode o meu cérebro compreender a si mesmo? Se meus *qualia* são estados do meu cérebro, como pode o meu cérebro descrever seu próprio funcionamento? Se enveredarmos por esse caminho, acabamos entrando num beco sem saída.

Devemos concluir então pelo abandono da neurociência? É óbvio que não. O que deve ser abandonada é a esperança de que a neurociência se tornará uma panaceia para a filosofia da mente e para toda a ciência cognitiva. Não temos razões para supor que a filosofia da mente vá desaparecer com o desenvolvimento da neurociência. Ao contrário, a neurociência parece depender de uma teoria filosófica da mente, para que seus achados empíricos possam ser interpretados em termos de nossa atividade mental.

O QUE LER

RYLE, G. *The Concept of Mind*

SELLARS, W. *Empiricism and the Philosophy of Mind*

Sobre materialismo eliminativo:

CHURCHLAND, P.S. *Neurophilosophy*

5. Este argumento é uma variação de Haldane (1932), citado por Popper & Eccles (1977). Já nos referimos a ele no capítulo III.

CAPÍTULO VI

FUNCIONALISMO E MENTES ARTIFICIAIS

Suponhamos que um dia você queira aprender a jogar xadrez. Seu instrutor começa por mostrar um tabuleiro, explicar quais são as peças que compõem o jogo (rei, rainha, cavalo, bispo, peão etc.,) e que tipo de movimentos elas podem fazer nesse tabuleiro. Em seguida, o instrutor precisa explicar quais são os objetivos do jogo, ou seja, como efetuar jogadas com a finalidade de tomar o rei do adversário, caso contrário não se ganha a partida.

Na primeira etapa o instrutor explica as *regras* do jogo, na segunda, ele explica algumas *estratégias* para ganhá-lo. As regras me dizem o que posso fazer com as peças, as estratégias me mostram que ao fazer uma jogada, o que conta é a posição das peças em relação às de meu adversário. Em outras palavras, as regras estabelecem as funções (movimentos) das peças, mas essas funções só adquirem sentido na medida em que o jogo se desenvolve e os movimentos a serem feitos se definem a partir das jogadas do adversário. É preciso definir e redefinir o que fazer em seguida a partir dos movimentos executados pelo adversário; uma boa ou má jogada se define no contexto da partida.

Quando alguém quer me ensinar a jogar xadrez, normalmente usa um tabuleiro e peças com diferentes formatos, representando o rei, a rainha, o cavalo, o bispo etc. O tabuleiro pode ser de madeira ou de plástico, as peças podem ser de metal ou de marfim. Posso jogar xadrez sem ter um tabuleiro ou peças, utilizando apenas a tela de um computador, com um tabuleiro e peças virtuais que não posso apalpar com minhas mãos. Pouco importa o material de que é feito o tabuleiro ou as peças, o que me habilita a jogar xadrez é conhecer as regras e o objetivo do jogo. Se as peças e o tabuleiro forem de marfim, de nada me adiantaria saber a composição química desse material; tampouco a história biológica e evolucionária do elefante de onde vieram esses pedaços de marfim. Se todos os tabuleiros e peças dos jogos de xadrez que existem no mundo fo-

rem destruídos, ainda assim eu posso recriar o jogo com outro material, desde que eu saiba as regras, estratégias e objetivos do xadrez.

Isto quer dizer que o jogo de xadrez tem uma *realidade independente* do material que utilizamos para representar as peças e o tabuleiro. Mas não haveria jogo de xadrez se não dispuséssemos de algum material ou algum meio pelo qual possamos representar o tabuleiro, as peças e as suas regras. Em outras palavras, não podemos suprimir inteiramente o material com o qual construímos um tabuleiro e suas peças, mas podemos *variá-lo* quase que indefinidamente. Se meu tabuleiro e minhas peças forem de marfim, terei *um exemplo de jogo de xadrez executado com peças de marfim*, mas as regras e estratégias do xadrez não serão redutíveis ao marfim; há sempre a possibilidade de jogar com peças de madeira ou de plástico.

Imagine agora que você já aprendeu a jogar xadrez. Dê um passo a mais e imagine que você pode descrever o funcionamento mental da mesma maneira que um jogo de xadrez. Se você conseguir imaginar essa analogia, estará no caminho para se tornar um partidário do *funcionalismo*.

O funcionalismo é uma alternativa ao materialismo reducionista e ao dualismo. Seu ponto de partida consiste de três pressuposições básicas: a) a realidade dos estados mentais (existe algo como um jogo de xadrez, com peças, regras e estratégias); b) a ideia de que os estados mentais não são redutíveis a estados físicos (o tabuleiro e as peças podem ser de madeira, de plástico ou de marfim – pouco importa – o que conta são as funções das peças, as regras e as estratégias) e c) os estados mentais são definidos e caracterizados pelo *papel funcional* que ocupam no caminho entre o *input* e o *output* de um organismo ou sistema (as estratégias, no jogo de xadrez, e as jogadas boas ou más se definem no contexto do jogo). Esse papel funcional caracteriza-se seja pela interação de um estado mental com outros que estejam presentes no organismo ou sistema, seja pela interação com a produção de determinados comportamentos.

Mas por que essas três pressuposições tornam o funcionalismo uma doutrina neutra que não favorece nem o dualismo nem o materialismo? Além de você precisar aprender xadrez e aplicar essa analogia para o funcionamento mental, é preciso agora que você imagine que o tabuleiro e as peças sejam o cérebro e que as regras e as estratégias sejam a mente. Como dissemos acima, se meu tabuleiro e minhas peças forem de marfim terei um exemplo de jogo de xadrez executado com peças de marfim; da mesma maneira, se o tabuleiro e as peças forem estados cerebrais, terei um exemplo de jogo de xadrez executado com estados cerebrais. Mas se prosseguirmos com a analogia, veremos também que a mente (as regras do jogo e as estratégias) não se reduzem ao cérebro. Pouco adianta eu estudar a composição química específica do cérebro, ela não me dará a chave para que eu possa compreender o funcionamento mental, da mesma maneira que a composição química das peças de marfim num jogo de xadrez não me permite entender suas funções e as regras do jogo.

Chegamos assim a um materialismo não reducionista: a mente ocorre no cérebro, esse faz o papel do tabuleiro e das peças, mas não me dá as regras nem as estratégias do jogo. O cérebro instancia uma mente, mas essa não *é* o cérebro nem se *reduz* a ele. Por

outro lado, para jogar xadrez, precisamos de algum tipo de material que possa representar as peças; não é possível que o jogo seja uma pura abstração sem qualquer tipo de correspondência material. A partir dessa pressuposição, o funcionalista pode rejeitar o dualismo substancial e propor uma teoria neutra, nem materialista nem dualista. Na realidade, o funcionalismo é *compatível* com o dualismo e com o materialismo não reducionista.

O funcionalismo tem explorado preferencialmente a compatibilidade com o materialismo não reducionista, deixando de lado a alternativa dualista[1]. A possibilidade de variar o tipo de material com o qual podemos construir mentes tem dado ao funcionalismo um lugar de destaque na filosofia da mente nas últimas décadas. Essa possibilidade tem servido de fundamento para as realizações tecnológicas oriundas de disciplinas novas como é o caso da inteligência artificial. A noção de uma inteligência *artificial* como realização de tarefas inteligentes – ou seja, a possibilidade de replicação mecânica de segmentos da atividade mental humana – por dispositivos que não têm a mesma arquitetura nem a mesma composição biológica e físico-química do cérebro foi a grande motivação para o aparecimento das teorias funcionalistas.

O funcionalismo consiste, assim, num nível de descrição onde é possível *abster-se de* ou *suspender* as considerações acerca da natureza última do mental, isto é, se esse é ou não, em última análise, redutível a uma estrutura física específica. A descrição das funções é uma descrição abstrata, que tem o mesmo estatuto da descrição de um software ou fluxograma que estipula quais as instruções que um computador deve seguir para realizar uma determinada tarefa. Os estados mentais têm uma *realidade* que deve ser reconhecida: a mente é o software do cérebro; mas esse software poderia ser instanciado em algum outro tipo de mecanismo que pudesse rodá-lo. Um software estipula um conjunto de regras – regras lógicas – e o que faz com que um cérebro ou um computador trabalhe de acordo com essas regras ou leis lógicas não é uma propriedade física específica de cérebros ou de computadores. Cérebros e computadores se equivalem desde que possam rodar esse software. Dois computadores podem diferir fisicamente um do outro embora possam trabalhar de acordo com as mesmas leis da lógica ou instanciar o mesmo software. Inversamente, dois computadores podem ser idênticos do ponto de vista físico, mas realizar tarefas inteiramente distintas se seu software for diferente.

A mesma analogia vale para mentes e organismos: um mesmo papel funcional que caracteriza um determinado estado mental pode se instanciar em criaturas com sistemas nervosos completamente diferentes, e nesse caso diremos que eles estão no *mesmo* estado mental. Um marciano pode ter um sistema nervoso completamente diferente do meu, mas se o sistema nervoso desse marciano puder executar as mesmas *funções* que o meu, o marciano terá uma vida mental igual à minha. Da mesma maneira, a

1. Uma exceção são os trabalhos de Chalmers, dos quais falamos no capítulo IV.

explicação da vida mental não pode ser buscada apenas por uma descrição do substrato físico que a instancia: um aparelho de rádio (hardware) toca uma música (software): a música e o aparelho de rádio são coisas distintas, irredutíveis uma à outra, embora ambas sejam necessárias para que possamos ouvir uma música. Nunca poderemos descrever *o que* o rádio está tocando através do estudo das peças que o compõem, sejam elas válvulas ou transístores.

Assim, o funcionalismo reconhece a existência de um nível de descrição que se sobrepõe ao nível puramente físico – e é nesse nível que devemos procurar uma explicação da atividade mental. Em outras palavras, a proposta funcionalista para uma teoria da mente consiste em sustentar que essa deve se constituir fundamentalmente como uma *ciência dos estados mentais* e do modo como esses interagem uns com os outros e que essa interação seria análoga ao software de computador.

Essa proposta acaba se opondo a qualquer perspectiva reducionista se considerarmos, por exemplo, que a especificação das bases biológicas e químicas de um acesso de raiva não explicam *do que* temos raiva. Da mesma maneira, não podemos saber *do que* temos medo quando especificamos as bases biológicas de um ataque de pânico nem tampouco com *o que* alguém sonha quando olhamos seu eletroencefalograma e nele identificamos as fases de sono REM[2]. Os conteúdos mentais requerem um nível de explicação próprio, que escapa à redução física.

A mesma coisa se aplica aos computadores: podemos encontrar equivalentes físicos de partes de um software nos estados físicos de um computador num determinado momento. Quem conhece computação sabe que é possível reduzir um software à chamada linguagem de máquina, embora essa seja uma tarefa extremamente trabalhosa. Nessa redução encontraremos uma imensa sequência de 0s e 1s, correspondendo aos estados "on" (ligado) e "off" (desligado) pelos quais passa a máquina, ou seja, aos diversos estados do hardware num determinado momento. Contudo, se alguém quiser reconstruir o programa que está sendo rodado a partir da linguagem de máquina encontrará grandes dificuldades para fazê-lo, na medida em que poderá haver um ou mais softwares compatíveis com uma mesma linguagem de máquina. A dificuldade aumenta se imaginarmos dois computadores rodando linguagens diferentes e simulando duas situações incomparáveis (a guerra do Kuwait e um jogo de xadrez, por exemplo) embora exibindo a mesma linguagem de máquina após esse processo de redução. Os dois computadores estarão em estados de hardware equivalentes, mas não poderemos, a partir disto, inferir aquilo que estão rodando: a redução falha ou é insuficiente.

Esse é um dos grandes atrativos do funcionalismo: não termos de esperar por qualquer decisão acerca da natureza dos estados mentais (se eles são estados físicos ou não) para podermos iniciar uma ciência da mente. Mas o funcionalismo apresenta também algumas dificuldades metodológicas que não podem ser omitidas. Ao tentar

2. Sobre essa noção, cf. capítulo I, nota 2.

estabelecer uma ciência dos estados mentais deparamos imediatamente com uma limitação: somente seres dotados de linguagem podem comunicar-nos realmente o que estão pensando, ou seja, descrever seus estados mentais. Isto equivale a dizer: é preciso saber, de antemão, o software que está sendo rodado por um computador para podermos descrever o que ele está fazendo. Por outro lado, se não temos esse acesso linguístico, resta-nos apenas a *atribuição* de estados mentais a outras pessoas com base na observação de seus comportamentos. Essa atribuição pressupõe, por sua vez, a existência de algum tipo de ligação lógica entre comportamentos e estados mentais que aproxima o funcionalismo do behaviorismo. Essa dificuldade só poderia ser superada se conseguíssemos descobrir que tipo de programa computacional é rodado pelas mentes humanas – o que era esperado pelos pesquisadores da inteligência artificial da década de 1970. Esses pesquisadores achavam que todos os aspectos do comportamento humano poderiam ser formalizados e descritos na forma de um programa de computador. Mesmo assim, haveria ainda a dificuldade de dois tipos diferentes de estados mentais se manifestarem através de um mesmo comportamento (por exemplo, posso abrir a porta porque sinto calor ou porque quero que ela fique aberta).

Apesar dessas dificuldades metodológicas, o funcionalismo teve um grande triunfo na década de 1970. Em 1975, o filósofo norte americano Hilary Putnam publica um célebre artigo "Minds and Machines" [Mentes e máquinas] sobre as relações mente-cérebro concebidas do ponto de vista da inteligência artificial, defendendo também um tipo específico de funcionalismo, baseado numa analogia entre mentes, cérebros e computadores digitais. A ideia de Putnam é que o computador digital fornece uma excelente analogia ou um bom modelo para concebermos a relação mente-cérebro: de um lado, há um conjunto de regras abstratas (instruções) e de outro, a realização física dessas regras através dos diferentes estados da máquina. Assim, a analogia consiste basicamente em estabelecer uma correlação entre estados mentais (pensamentos) e o software (conjunto de instruções da máquina ou o programa do computador) de um lado, e entre estados cerebrais e o hardware ou os diferentes estados físicos pelos quais passa a máquina ao obedecer às instruções. Essa correlação dispensaria qualquer tipo de pressuposição metafísica responsável pela possibilidade de interação entre o físico e o mental.

A proposta dessa correlação não constitui uma solução para o problema da relação entre mente e cérebro, na medida em que, pelo exame dos estados do hardware de um computador, não poderíamos inferir que tipo de software esse estaria rodando (não nos esqueçamos de que um computador simulando a guerra do Iraque contra o Kuwait e um jogo de xadrez podem exibir a mesma linguagem de máquina após um pro cesso de compilação). Ou seja, sem uma interpretação prévia ou uma reconstrução *a posteriori* dos conteúdos mentais, a correlação entre software/hardware e mente/cérebro não seria possível. Da mesma maneira que o exame do eletroencefalograma de uma pessoa em sono REM não permite saber com o que ela está sonhando, o exame das variações do hardware de um computador não permite saber exatamente o que esse está processando. Nesse sentido, o modelo proposto por Putnam não traz nenhuma contribuição substancial para esclarecer a principal dificuldade enfrentada

pelo problema da relação entre mente e cérebro, qual seja, saber como se dá a passagem do físico para o mental.

A despeito das críticas que podemos formular ao modelo de Putnam não podemos, entretanto, esquecer alguns méritos do funcionalismo, na medida em que esse chama a atenção para fatores metodológicos importantes a serem considerados nas tentativas de se formular uma ciência da mente. O modelo explicativo funcionalista parece ter se inspirado nas tentativas de simulação computacional das linguagens naturais, que preencheram a agenda da inteligência artificial durante a década de 1970. Pensemos na produção da linguagem humana enquanto processo físico. Quando falamos emitimos sons. Esses sons certamente constituem algo físico, detectável e mensurável, mas quando falamos emitimos também significados linguísticos. Ora, estamos dispostos a reconhecer a existência desses últimos, embora não possamos medir nem detectar significados da mesma maneira que o fazemos com coisas físicas. Há certamente uma independência entre sons e significados, embora esses últimos, para poderem se manifestar, dependam da existência dos primeiros. Contudo, o significado não depende de nenhuma propriedade física especificável que poderíamos detectar nas ondas sonoras: precisamos reconhecer a necessidade de um outro nível de descrição desses processos, que não é obtido pela sua descrição em termos físicos[3].

Os teóricos da inteligência artificial dos anos de 1970 defenderam uma espécie de "funcionalismo encarnado", sustentando a compatibilidade entre as descrições funcionais e sistemas físicos que as instanciariam. Contudo, uma outra dificuldade pode se manifestar aqui: até que ponto *qualquer* sistema físico poderia produzir vida mental? Como poderíamos, por exemplo, sustentar que um robô (digamos o COG ou o Hall de "2001 – Uma odisseia no espaço" ou mesmo um organismo de outro planeta) *sente dor* se não pudermos relacionar causalmente seus comportamentos que manifestam dor com uma estrutura neurológica similar à do ser humano? Será que podemos falar de uma "dor marciana" ou de uma "dor sentida pelo robô" que seria funcionalmente equivalente (ou seja, produziria os mesmos tipos de comportamento), mas que não seria uma dor como a sentida pelos humanos? Essas são apenas algumas das dificuldades que os defensores do funcionalismo tiveram de enfrentar – dificuldades que parecem se sobrepor a seus méritos. Falaremos dessas objeções em seguida; antes, porém, precisamos entender o projeto filosófico geral de descrição da mente proposto pelos funcionalistas e detalhar como eles conceberam as relações entre mente e cérebro de forma inovadora.

A linguagem do pensamento e a teoria representacional da mente

Afirmamos que o funcionalismo consolidou-se a partir do desenvolvimento da inteligência artificial que reforçou a analogia entre mentes e computadores. Construir

3. Faço essa mesma observação em Teixeira (1996a).

essa analogia não significa, entretanto, sugerir que nós sejamos robôs rigidamente programados. Na verdade, os funcionalistas propõem que nossa mente é o software de nosso cérebro – um software que poderia ser rodado em outro tipo de substrato físico, como, por exemplo, um computador digital. Essa possibilidade de rodar um mesmo software em diferentes tipos de hardware ou de substrato físico é chamada de tese da *múltipla instanciação* (*multiple realizability*) – uma tese que tem um lugar central na proposta funcionalista. Em outras palavras, se dois sistemas forem capazes de executar as mesmas funções, eles são equivalentes.

Essa tese não significa que o material com o qual os sistemas são construídos deva ser ignorado ou seja irrelevante, mas apenas que pode ser abstraído quando tentamos explicar suas funções, da mesma maneira que abstraímos propriedades específicas de um polígono qualquer quando queremos efetuar uma demonstração geométrica. Nesse caso, abstraímos a cor e o material de que é feito o polígono e consideramos apenas suas propriedades geométricas. Da mesma maneira, o funcionalismo abstrai o material no qual é instanciado um determinado software, considerando apenas suas propriedades funcionais. Em outras palavras, optamos por uma determinada perspectiva ou um determinado *nível de explicação*, que, no caso, corresponde à descrição da atividade do software da mente, de onde se derivaria uma ciência do pensamento ou uma ciência dos estados mentais e do modo como esses são organizados. Dessa ciência não deve participar uma explicação do modo pelo qual esses estados podem ser produzidos, ou seja, uma ciência da mente não deve se preocupar com o substrato físico que leva à produção de estados mentais. A explicação do modo como estados mentais podem ser produzidos caberia à neurociência – caso esses sejam produzidos pelo nosso cérebro – ou à engenharia, caso esses sejam produzidos pelo hardware de um computador. Aplica-se aqui a tese da múltipla instanciação, ou seja a ideia de que cérebros e computadores se equivalem na medida em que possam desempenhar um mesmo conjunto de funções – funções descritíveis como softwares e que levariam à produção da vida mental.

Ora, como seria possível construir uma ciência da mente considerando apenas os estados mentais e suas possíveis relações? O ponto de partida é a ideia de que estados mentais são, essencialmente, proposições. Proposições são expressas por sentenças da nossa linguagem, mas não são a mesma coisa que essas. Considere, por exemplo, as duas sentenças: "Está chovendo" e "*It is raining*". Temos duas sentenças diferentes, em línguas diferentes, embora elas expressem a mesma proposição, ou seja, o fato de que está chovendo. A ciência da mente na acepção funcionalista parte da ideia de que o pensamento é um conjunto de proposições que se tornam sentenças na nossa cabeça. É preciso estudar o modo como essas proposições se relacionam, ou seja, o conjunto de regras subjacente a sua organização: isto nos permitirá entender a própria estrutura de nosso pensamento.

Esse estudo pode ser feito se assumirmos que proposições (e as sentenças nas quais elas são expressas) são símbolos. Pensar é, então, manipular símbolos, e a ciência da mente deve ser o estudo das regras e relações que coordenam essa manipulação simbólica. Isto

nos traz de volta à analogia entre mentes e computadores, ou melhor, entre mentes e programas computacionais. Programas computacionais estipulam relações entre símbolos e as regras envolvidas nessas. Ora, o estudo das relações entre símbolos nada mais é do que aquilo que chamamos de *sintaxe*. A mente passa então a ser vista como uma grande e complexa máquina sintática que, a partir de computações, produz aquilo que chamamos de pensamento. O software da mente está escrito em símbolos e sua sintaxe governa a sequência de produção de pensamentos.

Por uma analogia com nossa própria linguagem – também um conjunto de símbolos governados por regras sintáticas – podemos chamar o software que constitui nossas mentes de *linguagem do pensamento*. Esse é o nome que lhe conferiu J. Fodor em seu livro clássico *The Language of Thought* [A linguagem do pensamento], publicado em 1975. A imagem da mente como uma máquina sintática baseia-se na ideia de que os pensamentos ou símbolos são "representações mentais". É por essa razão que a concepção desenvolvida por Fodor passou a ser chamada também de Teoria Representacional da Mente ou RTM (*Representational Theory of Mind*). A tese central da RTM é que a mente manipula representações mentais na forma de sentenças não interpretadas que se encontram na linguagem do pensamento.

Ora, como isto é possível, ou seja, que nossa mente manipule sentenças mesmo sem que essas sejam interpretadas ou tenham um significado? Seria perfeitamente plausível imaginar um dispositivo que manipulasse sentenças independentemente de saber seu significado – ou seja, dispositivo que, embora não sabendo o significado das sentenças, as manipulasse *como se* o soubesse. Esse dispositivo corresponderia a algo que imita o desempenho linguístico de um falante nativo de uma língua qualquer – um imitador que não saberia o significado das sentenças que estaria produzindo e manipulando. Esse imitador basear-se-ia apenas no "formato" dos símbolos que produz e manipula, ou seja, com base apenas na sua sintaxe ou posição relativa no interior de um sistema simbólico amplo.

Ora, que tipo de dispositivo poderia ser esse? Certamente trata-se de um computador digital – uma máquina que funciona unicamente a partir de regras sintáticas e princípios formais. Essas regras sintáticas e princípios formais de organização dos símbolos são a base para a produção de sentenças dotadas de significado, ou seja, a semântica é construída a partir da sintaxe. Um computador digital poderia, apenas a partir de um conjunto de símbolos e regras sintáticas, produzir sentenças com significado. Nossas mentes seriam esse tipo de computador. Máquinas que processam sintaticamente representações, sem que essas precisem ser previamente interpretadas para que possam adquirir significado. Desvendar as regras sintáticas a partir das quais nossa mente processa representações é, assim, a tarefa prioritária de uma ciência da mente. Pouco importa se essas representações são processadas por cérebros ou por computadores, desde que esses sejam funcionalmente equivalentes.

Se examinarmos um cérebro, certamente não encontraremos nele nada parecido com símbolos. Contudo, isto não é importante. Todos nós concordamos com a ideia de que computadores processam símbolos, apesar do fato de que, se abrirmos um compu-

tador, não encontraremos nele nada que se pareça com um símbolo, nem mesmo os "0s" e "1s" que compõem a linguagem de máquina. Os componentes internos de um computador não se assemelham em nada ao modo como representamos símbolos no papel, da mesma maneira que a banda magnética de um CD em nada se assemelha à música que esse produz quando o tocamos. Se operações mentais envolvem a manipulação de símbolos na linguagem do pensamento, o modo como esses são processados pelo cérebro em nada se assemelha ao modo como os representamos no papel. Símbolos podem ser instanciados através da atividade eletroquímica do cérebro ou mesmo nas conexões sinápticas entre os neurônios, mas nós nunca os perceberíamos se abríssemos um cérebro.

Mas há ainda uma questão a ser respondida pelos proponentes da hipótese da linguagem do pensamento: como os símbolos que a compõem adquirem significado? A resposta proposta por Fodor é: através da relação do cérebro com o meio ambiente. O significado de um símbolo não lhe é intrínseco e depende da relação causal entre organismo e meio ambiente. Ou seja, o significado dos termos da linguagem do pensamento não é fixo, variando na medida em que o meio ambiente no qual se situa o organismo também variar. O termo "abacaxi" pode designar abacaxis se, por exemplo, percebermos abacaxis a nossa volta ao caminhar por uma feira. Em outras palavras, a semântica da linguagem do pensamento é determinada pelo contexto no qual o organismo se encontra.

Na medida em que o significado dos termos da linguagem do pensamento é determinado contextualmente, seu papel é secundário na elaboração de uma ciência da mente. Essa deve se preocupar primordialmente com as relações sintáticas entre os símbolos, pouco importando a que eles se referem no mundo, pois isto requereria que a ciência da mente incorporasse o estudo dos ambientes nos quais o organismo está situado ou pode situar-se. Isto requereria, por sua vez, que a ciência da mente extrapolasse seu âmbito e seus objetivos, ou, para usar as próprias palavras de Fodor, "que a psicologia herdasse o mundo" (FODOR, 1981). Um exame das relações sintáticas é tudo o que a ciência da mente deve fazer; um exame que deve isolar e abstrair, pelo menos como estratégia metodológica, a relação dos termos da linguagem do pensamento com seus referentes no mundo. Esse isolamento ou abstração metodológicos seriam necessários para tornar a ciência da mente exequível e constitui aquilo que Fodor chamou de "solipsismo metodológico" (FODOR, 1981).

A proposta de Fodor engloba também uma maneira peculiar de conceber as relações entre mente e cérebro que se tornou ortodoxa entre os defensores do funcionalismo. Ao enfatizar a prioridade das relações sintáticas entre símbolos e entre as sentenças da linguagem do pensamento, Fodor dá um passo inovador. Símbolos não dependem de intérpretes e suas propriedades sintáticas derivam-se do fato de eles constituírem algo *físico*. Isto torna as sentenças da linguagem do pensamento entidades concretas que desempenham um papel causal na determinação da cognição e do comportamento. A relação entre mente e cérebro é análoga àquela de um programa de computador e sua implementação em vários tipos de hardware. Mas a ideia de que símbolos se-

jam algo físico e que as sentenças da linguagem do pensamento sejam algo concreto dá ao funcionalismo de Fodor um caráter peculiar ao torná-lo compatível com a hipótese materialista. Fenômenos mentais não são redutíveis nem idênticos a fenômenos cerebrais – apesar de serem algo físico – da mesma maneira que programas de computador não são idênticos nem redutíveis a nenhum tipo de hardware. É possível falar de mentes como algo diferente do cérebro sem, no entanto, incorrer no dualismo ou na tese da imaterialidade dos fenômenos mentais. O materialismo não reducionista proposto por Fodor afasta-o, igualmente, das teorias da identidade entre mente e cérebro que analisamos no capítulo III.

As objeções ao funcionalismo

O funcionalismo tem suscitado uma série de debates filosóficos nos quais, às vezes, se evidenciam alguns de seus aspectos paradoxais. Podemos agrupar as principais objeções feitas ao funcionalismo na literatura contemporânea de acordo com os seguintes critérios: a) o funcionalismo desconsidera a importância da base cerebral e biológica que seria responsável pela produção daquilo que chamamos mente, b) a descrição funcional da mente entendida como um computador não é capaz de apreender uma das características fundamentais dos estados mentais, qual seja, sua natureza significativa ou semântica (esse é o famoso argumento do quarto do chinês, formulado por John Searle em 1980a) e c) o funcionalismo não apreende uma das características centrais da mente, qual seja, os *qualia*. Veremos que todas essas objeções são, entretanto, insuficientes para rejeitar o funcionalismo – se esse for entendido de uma forma mais ampla do que simplesmente uma identificação entre mentes e programas de computador.

Funcionalismo e cérebro – quando se afirma que o cérebro é um computador e que a mente é o seu software enfrentamos desde logo uma dificuldade: o que devemos entender por um computador? Essa pode parecer uma questão trivial ou até ingênua, mas não é isto que ocorre do ponto de vista filosófico. Será que qualquer dispositivo físico que execute um conjunto de regras ordenadas de uma maneira determinada (um algoritmo) deve ser considerado um computador? Em outras palavras, até que ponto podemos estender a teoria da múltipla instanciação (*multiple realizability*) sem incorrer em paradoxos? O funcionalismo toma como ponto de partida a ideia de que a matéria física sobre a qual é implementado um programa ou a execução de um conjunto de regras ordenadas pode ser variado, desde que se preserve sua adequação funcional.

Podemos caricaturar isto e dizer que, para o funcionalista, um gato e uma ratoeira são a mesma coisa se a característica funcional de ambos for apanhar ratos. Sem incorrer nesse exagero, podemos, contudo, construir um dispositivo qualquer usando pedaços de arame e tampinhas de refrigerante e dizer que esse dispositivo é um computador se ele executar funções que possam ser descritas ao modo de uma máquina de Turing, isto é, se esse dispositivo estrambótico for capaz de seguir um conjunto de regras, passo a passo, que tenham sido estipuladas para se chegar a um estado final (output) a par-

tir de um estado inicial (input)[4]. Qualquer coisa na natureza poderia ser considerada, no limite, um computador, se executasse computações que levassem de um estado inicial a um estado final. Desse ponto de vista, a própria noção de computador torna-se excessivamente genérica.

Meu dispositivo estrambótico pode ser manipulado com as minhas mãos – eu posso trocar tampinhas de refrigerante de um lugar para o outro usando apenas minhas mãos. Computadores mais antigos, construídos na década de 1940, exigiam que várias operações de conectar fios em soquetes etc. fossem executadas manualmente. Intuitivamente não consideraríamos mais, hoje em dia, tais máquinas como sendo computadores, muito menos meu dispositivo estrambótico feito de tampinhas de refrigerante. Contudo, o que é mais paradoxal nessa história é o fato de até hoje não dispormos de uma definição unívoca do que seja um computador. Ao se definir a ideia de computação de modo matemático, ou seja, de modo essencialmente abstrato, amplia-se excessivamente o conjunto de dispositivos que podem ser considerados um computador e essa noção, ao tornar-se excessivamente genérica pode igualmente tornar-se absolutamente trivial. Definir computação abstratamente trouxe um ganho em termos de múltipla instanciação e daquilo que se chama, em ciência da computação, de *portabilidade*. A portabilidade significa definir um programa em termos de um conjunto de relações sintáticas entre símbolos, o que permite que um mesmo programa seja rodado em hardwares completamente diferentes. A ideia é tentadora e útil, pois se inteligência é programa e se programas são portáveis e independentes dos substratos materiais onde podem ser realizados, abre-se a perspectiva de poder replicar a inteligência sem ter de replicar o cérebro, ou seja, a própria possibilidade de se construir inteligências *artificiais*.

Problemas podem surgir, contudo, quando se parte da ideia abstrata de computação, de portabilidade e de múltipla instanciação para se estabelecer uma analogia entre computadores e funcionamento mental. Esses três princípios, por serem excessivamente genéricos, podem levar a paradoxos. Intuitivamente eu me recusaria a dizer que meu dispositivo estrambótico de tampinhas de garrafa de refrigerante poderia ser considerado a simulação de qualquer aspecto da mente humana. Mas posso estender ainda mais essa ideia a ponto de torná-la ridícula.

O funcionalismo ganhou espaço na década de 1970 quando se tinha ainda resquícios da ideia de que o cérebro poderia ser modelado como uma máquina digital, ou seja, que os neurônios e as conexões entre neurônios funcionariam ao modo de "ligado" (*on*) ou "desligado" (*off*) ou em termos de "sim" e não". Sabe-se hoje em dia que é possível que algumas partes do cérebro funcionem dessa maneira, mas não todas. Debate-se até hoje se a passagem de impulso entre um neurônio e outro, o potencial de ação, é uma questão de "tudo" ou "nada". Sabe-se que entre um disparo e outro de um

4. Para uma descrição mais detalhada da ideia de uma máquina de Turing, cf. Teixeira (1998), capítulo I.

neurônio há gradações que nos inclinariam a supor que o cérebro seria uma máquina analógica e não digital. Contudo, nos anos de 1970 ainda prevalecia a ideia do cérebro entendido como uma máquina predominantemente digital.

Ora, se o que conta são as funções e, se essas podem ser reproduzidas por um dispositivo digital que opera em termos de "sim" e não", pouco importando seu substrato físico desde que ocorra a transmissão de impulso elétrico, pode-se concluir não apenas que mentes não requerem cérebros como também que elas podem *prescindir totalmente* desses últimos. Suponhamos que eu queira modelar um cérebro dotado de um bilhão de neurônios. Eu poderia fazer isto pedindo para cada habitante da China efetuar o que um neurônio faz, ou seja, poderia pedir para cada um deles sinalizar a passagem ou não de corrente elétrica usando bandeirinhas. Quando a bandeirinha fosse levantada, isto quereria dizer "sim" (*on*) e quando estivesse abaixada, quereria dizer "não" (*off*). A população da China estaria implementando computações e, se implementasse computações independente de um substrato material específico, eu poderia, como bom funcionalista, afirmar que a população da China com suas bandeirinhas é uma simulação não só de um cérebro como também de uma mente!

Esse paradoxo foi notado pela primeira vez pelo filósofo Ned Block em 1978. Poderíamos explorá-lo ainda mais, levando-o a um outro extremo. Eu poderia construir uma simulação do cérebro de alguém, utilizando-me de neurônios artificiais que funcionariam como portas do tipo "sim" e não". Mas, em vez de fazer uma fiação específica ou um circuito que comunicasse esses neurônios uns com os outros (uma imitação das sinapses), eu poderia dotar, cada um, de um microtransmissor de ondas de rádio que, como aquelas ondas utilizadas para guiarmos nossos aeromodelos, emitisse sinais entre esses neurônios, comunicando quando eles deveriam estar na posição "sim" (porta aberta) ou na posição "não" (porta fechada). Meu sistema de transmissão de ondas, cuidadosamente instalado em cada um dos neurônios, poderia ser particularmente poderoso, viajando longas distâncias. O passo seguinte seria então espalhar esses neurônios pelo universo, situando cada um em um planeta de uma galáxia distante. Quando ativados, eles funcionariam como um cérebro e estariam também replicando uma mente espalhada pelo cosmos. A conclusão paradoxal que chegamos, então, é que, do ponto de vista do funcionalismo, não apenas podemos prescindir de cérebros para replicar mentes, como também que podemos prescindir de qualquer tipo de característica neuroanatômica e neurofisiológica que estaria presente em nossos cérebros. Desenvolver uma ciência do cérebro seria uma tarefa inglória e inútil, algo totalmente desnecessário para compreender o funcionamento da mente humana[5].

Esse tipo de objeção enfatiza o contraste entre a proposta funcionalista e os avanços recentes da neurociência, na medida em que essa caminha na direção de mostrar que o cérebro humano se assemelha muito mais a uma máquina eletroquímica do que a

5. Sobre esse tipo de paradoxo cf. o conto de Zuboff, "The Story of a Brain", In: Dennett & Hofstadter (1981).

uma placa de computador. Vimos no capítulo III que as ligações entre neurônios que ocorrem no interior do cérebro são efetuadas a partir de neurotransmissores, mensageiros e receptores, que atuam a partir de sua estrutura química. Ligações entre neurônios dependem da produção desses neurotransmissores e ocorrem na medida em que esses neurotransmissores encontram receptores num esquema do tipo "chave e fechadura" (*lock and key*). Só assim a informação, na forma de impulso elétrico é transmitida entre os vários neurônios. A transmissão depende da produção dessas substâncias químicas e é operada através delas. Se, por um lado, isto torna o processamento de informação no cérebro mais lento, por outro, permite que as ligações sejam feitas e desfeitas, ou que novas ligações possam surgir a qualquer momento, dotando o cérebro de uma grande plasticidade que lhe permite reagir à variação das circunstâncias ambientais. Ora, esse é um modelo bastante distante do hardware de um computador, onde todas as possíveis ligações já estão preestabelecidas entre seus chips.

Ainda seguindo essa linha de raciocínio, um outro conjunto de objeções ao funcionalismo foi levantado no final da década de 1970. Essas objeções surgiram a partir do paradoxo formulado por Block. Por que a ideia de que os habitantes da China poderiam simular uma mente parece-nos tão contra intuitiva? A resposta a que se queria chegar era: a organização funcional é insuficiente para explicar a produção de fenômenos mentais. Não poderíamos prescindir da natureza do substrato físico no qual a mente seria instanciada. Além da organização funcional, precisaríamos também levar em conta características específicas do cérebro, como, por exemplo, sua estrutura bioquímica. Sem essas características específicas do substrato cerebral não poderíamos explicar a natureza dos *qualia* e da experiência consciente.

Esses argumentos contrários ao funcionalismo tomaram como ponto de partida a comparação entre sistemas funcionalmente equivalentes. É inegável, por exemplo, que eu tenha experiências conscientes e *qualia*. Mas será que isto ocorreria necessariamente num sistema, simplesmente por esse ser funcionalmente idêntico a mim? Block foi o primeiro a apontar para essa dificuldade ao formular o "argumento dos *absent qualia*" ou *qualia* ausentes. O "cérebro" de um robô pode ser funcionalmente igual ao meu, mas isto não implica que esse cérebro artificial possa ter sensações subjetivas iguais às minhas ou experiências conscientes. Estar nos mesmos estados funcionais seria condição necessária, mas não suficiente, para a produção dessas sensações subjetivas. E se essas não ocorrem no robô, isto se deve, muito provavelmente, ao fato de o cérebro desse ser de um material diferente do cérebro de um ser humano. Um robô funcionalmente equivalente a mim seria, muito provavelmente, um zumbi sem experiências conscientes. Mas mesmo que não o fosse, isto revelaria ainda um outro fato: a descrição funcional seria incompleta, isto é, ela não me permitiria saber se o robô – ou um sistema funcionalmente equivalente a mim – experiencia ou não *qualia* e estados mentais conscientes.

Uma exploração mais sofisticada desse tipo de argumento encontramos em Shoemaker (1975). A objeção de Shoemaker ou "argumento dos *inverted qualia*" (*qualia*

invertidos) sustenta que podemos imaginar que existem criaturas funcionalmente equivalentes a nós que, em vez de experienciar cores da nossa maneira, têm experiências completamente diferentes. Quando enxergamos verde, ao perceber uma árvore, esse tipo de criatura experienciará o vermelho – e isto se deveria, supostamente, a diferenças na composição de seu substrato cerebral. Contudo, seus comportamentos e seus estados internos (funcionais) seriam indistinguíveis do nosso.

Ambos os argumentos – de Block e de Shoemaker – estariam apontando para o fato do funcionalismo ser uma teoria psicológica inadequada, por ser excessivamente genérica; incapaz de detectar diferenças subjetivas ou *qualia*. Esses argumentos – verdadeiras máquinas de guerra contra o funcionalismo – suscitaram um debate interminável na filosofia da mente, que se estende até hoje. Voltaremos a falar deles mais adiante.

Funcionalismo e significado[6] – A discussão acerca da plausibilidade do funcionalismo toma um impulso decisivo a partir da publicação do artigo de J. Searle, "Minds, Brains and Programs" [Mentes, cérebros e programas][7] em 1980. Nesse artigo clássico, Searle apresenta uma crítica vigorosa à possibilidade de se obter um equivalente mecânico para o fenômeno cognitivo humano que normalmente denominamos *compreensão*. Seu ponto de partida foi a análise dos programas para compreender histórias curtas desenvolvidas por R. Schank na Universidade de Yale e que se concretizaram no trabalho *Scripts, Plans, Goals and Understanding* [Roteiros, planos, metas e compreensão] (1977). Por exemplo, ao fornecer a um computador a história de um homem que entra num restaurante, pede um sanduíche e sai sem pagar ou deixar gorjeta porque notou que o pão estava estragado, o programa de Schank era construído de tal maneira que lhe era possível responder coerentemente a questões elaboradas com base no texto da história. Tendo em vista esses resultados, Schank sustentava que seu programa era capaz de *compreender* o texto e constituía uma explicação para a capacidade do ser humano de compreender textos ou histórias curtas.

As críticas desenvolvidas por Searle às pretensões de que um tal programa realmente compreendesse histórias basearam-se na construção de um experimento mental que reproduzia num ser humano o procedimento do próprio programa de Schank. Imagine um falante trancado num quarto. Esse falante só conhece o português e tem em seu poder: a) um texto escrito em chinês, que pode, por exemplo, ser uma história; b) um conjunto de regras de transformação, em português, que permite executar operações sobre o texto em chinês. Essas operações são idênticas àquelas desempenhadas pelos programas de Schank: trata-se de operações de decomposição e recomposição de palavras com base num *script* que permite à máquina reconhecer palavras-chave em uma determinada questão, comparar a presença dessas palavras no texto e o de-

6. Parte do material desta seção encontra-se em Teixeira (1998).

7. Há tradução deste artigo para o português em Teixeira (1996b).

compor a partir dessas informações. O *script* contém informação relevante (por exemplo, sobre como são os restaurantes em geral, qual o procedimento para pedir comida etc.), o que capacita o sistema a responder às questões desejadas.

O falante (trancado no quarto) recebe periodicamente novos textos e questões em chinês e aplica essas regras de transformação associando as sequências anteriores com as sequências mais recentes. Com base nessas regras de transformação, passa a emitir ou escrever mais sequências de símbolos em chinês. Claro que o falante preso no quarto não sabe precisamente o conteúdo das informações que ele está gerando com base nos dois textos e nas regras de transformação. O primeiro texto corresponde, em nosso experimento mental, ao relato que é fornecido ao computador, o segundo texto ao conjunto de questões que é elaborado com base nesse relato; e as novas sequências geradas, às respostas a tais questões. As regras de transformação são bastante complexas, e concebidas de maneira tal que elas simulem os processos mentais e o comportamento linguístico de um falante nativo de chinês numa conversação habitual. Após um certo tempo o falante aprendeu a manipular perfeitamente essas regras de transformação e, com base nos outputs, um observador externo poderia dizer que ele *compreende* chinês – o que, no entender de Searle, constitui um contrassenso.

A instanciação dos programas de Schank num sujeito humano, reproduzida nesse experimento mental, é, para Searle, bastante reveladora. Ela mostra que os programas desse tipo não estabelecem as condições necessárias para a simulação da atividade cognitiva da *compreensão*: o falante aplica as regras de transformação e compreende essas regras, mas as sequências de símbolos em chinês não têm nenhum significado para ele. A manipulação de símbolos realizada no programa é inteiramente cega – uma manipulação de símbolos formais que não tem intencionalidade; "não é sequer manipulação de símbolos, uma vez que esses símbolos não simbolizam nada, ou seja, eles têm apenas sintaxe, mas não semântica"[8].

A esse programa falta intencionalidade intrínseca ou genuínos estados mentais. A atribuição de intencionalidade ou de *significado* diz Searle, é, nesses casos, sempre uma atribuição *a posteriori*, dependente da intencionalidade intrínseca dos sujeitos humanos que observam os outputs do programa.

Mas o que é "intencionalidade intrínseca" no entender de Searle, e que parentesco tem essa noção com a ideia de significado? O conceito de intencionalidade intrínseca não é muito explorado em "Minds, Brains and Programs". Searle o desenvolve com maior profundidade em outros artigos, tais como "Intrinsic Intentionality" [Intencionalidade intrínseca] (1980b) e "What is an Intentional State" [O que é um estado intencional] (1982), bem como no seu livro de 1983, *Intentionality* [Intencionalidade]. A intencionalidade, segundo Searle, é uma "capacidade" apresentada pelos seres vivos, através da qual nossos estados mentais se relacionam com os obje-

8. Cf. Searle (1981, p. 300s.).

tos e estados de coisas no mundo. Assim, se tenho uma intenção, essa deve ser a intenção de fazer alguma coisa, se tenho um desejo ou um medo, esse desejo e esse medo devem ser um desejo ou medo de alguma coisa que está no mundo. Um estado intencional pode ser definido, grosso modo, como uma representação associada a um determinado estado psicológico.

Essa mesma capacidade – estritamente biológica no entender de Searle – percorre nossa linguagem, convertendo-a num tipo particular de relação organismo/mundo. Contudo, ela não é uma propriedade da linguagem e sim uma propriedade específica que nossos estados mentais imprimem ao nosso discurso. Nessa operação, os sinais linguísticos, sejam eles os sons que emitimos ou as marcas que fazemos no papel, passam a ser representações de coisas ou estados de coisas que ocorrem no mundo, e no caso específico das representações linguísticas podemos afirmar que elas constituem descrições dessas representações ou mesmo representações de representações que estão na nossa mente. A intencionalidade dos estados mentais não é derivada de formas mais primárias da intencionalidade, mas é algo intrínseco aos próprios estados mentais. Nesse sentido, a intencionalidade é a propriedade constitutiva do mental e sua base é estritamente biológica – só os organismos desempenham essa atividade relacional com o mundo, constituindo representações. Sua origem está nas próprias operações do cérebro e na sua estrutura, constituindo parte do sistema biológico humano, assim como a circulação do sangue e a digestão.

A intencionalidade intrínseca, presente no discurso linguístico, constitui uma forma derivada de intencionalidade que consiste na relação das representações linguísticas com os estados intencionais, o que permite que estas últimas sejam representações de alguma coisa do meio ambiente. Em outras palavras, essa relação entre representações linguísticas e estados intencionais transforma o código linguístico num conjunto de signos, ou seja, estabelece o seu *significado*. Nesse sentido, a intencionalidade intrínseca constitui para Searle a condição necessária para que um sistema simbólico adquira uma dimensão *semântica*. Sem essa dimensão semântica não podemos falar de compreensão. E sem essa relação entre representações mentais ou estados intencionais e representações linguísticas não podemos falar de compreensão de textos ou compreensão linguística.

*Algumas respostas

Que tipo de respostas os partidários do funcionalismo poderiam dar a essas objeções? Procedendo em ordem inversa, analisaremos uma resposta aos argumentos de Block e de Shoemaker, para, em seguida, falarmos do argumento do quarto do chinês de Searle.

Uma resposta convincente aos argumentos dos *qualia* ausentes e dos *qualia* invertidos encontramos em Chalmers (1996a). Ele propõe dois tipos de experimentos mentais para mostrar que as situações imaginadas por Block e Shoemaker não poderiam acontecer: se se reproduz a organização funcional do cérebro, com essa se reproduz também a geração de *qualia* e de experiências conscientes. Os dois experimentos mentais propostos por Chalmers procedem por redução ao absurdo.

No primeiro experimento mental, imagina-se que os neurônios de alguém, por estarem em processo de deterioração devido a uma moléstia qualquer, vão sendo gradualmente substituídos por chips de silício. Num primeiro momento, opera-se a substituição de um único neurônio, preservando, contudo, a organização funcional do cérebro. Ou seja, o chip é conectado exatamente no lugar onde o neurônio estava e todas as conexões com outros neurônios são refeitas e preservadas. O chip preserva também todas as propriedades químicas e elétricas do neurônio substituído, simulando-as com rigorosa perfeição. Nesse sentido, a substituição não faz nenhuma diferença para o sistema como um todo.

Em seguida, substitui-se um segundo neurônio vizinho ao primeiro. E assim por diante, até que todos os neurônios dessa pessoa sejam substituídos por chips de silício. Esses chips fazem tudo o que os neurônios fariam, mas através de sinais elétricos e não por reações bioquímicas como ocorria antes no cérebro dessa pessoa. A partir daí passa-se a dizer que essa pessoa tornou-se uma outra pessoa ou algum tipo de robô – o robô do argumento dos *absent qualia* ou "*qualia* ausente" formulado por Block. Seu cérebro de silício seria funcionalmente equivalente ao cérebro natural de que ela estava dotada anteriormente e, se Block estiver certo, essa pessoa de cérebro de silício não terá experiências subjetivas ou *qualia*.

Ora, Block não considerou o que ocorreria nos estágios intermediários existentes entre o cérebro dessa pessoa – capaz de ter experiências subjetivas – e o cérebro do robô, feito de silício e no qual supostamente essas experiências deveriam cessar. O que ocorreria com a experiência subjetiva dessa pessoa nesses estágios intermediários, ou seja, quando seu cérebro tivesse dez por cento de chips de silício, ou 25%? Para nossa comodidade, chamemos a pessoa de cérebro natural de Hipólito e essa mesma pessoa, quando tem o cérebro totalmente substituído por chips de silício, de Atílio.

Essa é a alegoria que Chalmers nos conta. Hipólito vai assistir a um jogo de futebol. Senta-se no meio da torcida e começa a vibrar com seu time preferido, sempre olhando para as cores das camisas dos jogadores, que são, digamos, uma combinação de vermelho e amarelo. Mas se a pessoa que estivesse sentada lá na arquibancada fosse Atílio, ou seja, alguém com 25 ou 30% de seu cérebro substituído por chips, possivelmente não experienciaria as cores das camisas com a mesma intensidade de Hipólito. Como o cérebro de Atílio está numa situação transitória no que se refere à substituição de neurônios por chips, suas experiências subjetivas estariam, também, gradativamente diminuindo. O que será que Atílio veria em termos de cores? Talvez cores menos intensas, em vez de vermelho e amarelo, algo como rosa e bege. A experiência de cor tenderia a ir desaparecendo para Atílio na medida em que todos os seus neurônios fossem progressivamente sendo substituídos. Isto seria algo desagradável, embora, por outro lado, poderia haver vantagens. Quando Hipólito sente terríveis dores de coluna, Atílio sente apenas algum desconforto muscular – pois nesse caso também estaríamos numa situação de substituição parcial.

Chalmers argumenta que, se Block estivesse certo, chegaríamos a uma situação absurda. Atílio continuaria dizendo que ele estaria experienciando amarelo e verme-

lho ou uma terrível dor na coluna, quando, de fato, sua experiência seria de rosa e de bege ou de apenas uma pequena dor muscular. Pois, apesar de ter parte de seus neurônios substituídos por chips de silício, o cérebro de Atílio continua funcionalmente equivalente ao cérebro de Hipólito – ou seja, ao cérebro original, antes da substituição progressiva ter sido iniciada. Do ponto de vista funcional, ele teria de estar experienciando amarelo e vermelho, embora não seja isto o que ocorre. Das duas uma: ou Atílio está errado acerca daquilo que ele mesmo está percebendo ou seu cérebro não está no mesmo estado funcional que o cérebro de Hipólito quando esse percebe amarelo e vermelho.

Essa segunda possibilidade deve ser descartada imediatamente, pois a substituição foi feita cuidadosamente, ou seja, os chips de silício foram instalados de modo a funcionarem exatamente como os neurônios. Ademais, a equivalência funcional é a premissa principal sobre a qual se assenta o argumento dos *absent qualia*: se abríssemos mão dessa, solaparíamos o argumento de Block e esse perderia sua razão de ser. Restaria então a primeira possibilidade: Atílio teria de estar enganado acerca de suas próprias percepções. Mas isto seria um absurdo: uma pessoa não pode estar errada acerca daquilo que está experienciando. Imagine o que aconteceria se você estivesse olhando para algo bege e alguém te dissesse: "Não, você não está vendo nada bege. Aliás, embora você perceba algo bege, você *deveria* estar vendo algo amarelo".

Ora, esse argumento formulado por Chalmers refuta a possibilidade de *qualia* ausente ou *absent qualia*, mas ainda não diz nada acerca dos *inverted qualia* de Shoemaker. Shoemaker diria que Atílio poderia continuar a ter experiências da mesma maneira que Hipólito as tinha, mas que essas experiências poderiam ser radicalmente diferentes em termos de conteúdo, sem, entretanto, alterar o comportamento de Atílio. Um mesmo estado funcional poderia ocorrer em Hipólito e em Atílio, mas quando Hipólito estivesse experienciando vermelho Atílio estaria experienciando verde. Pensemos, por exemplo, que os dois dirigem carro. Hipólito pararia na frente de um sinal vermelho, e Atílio também, embora estivesse experienciando verde. Ocorre que para Atílio formou-se um *inverted qualia* ou seja, ele para o carro diante de um semáforo verde e não para diante de um semáforo vermelho. Como vermelho e verde foram coerente e sistematicamente invertidos nas experiências de Atílio, ambos podem dirigir com a mesma segurança.

Vamos agora introduzir mais um complicador nessa história. Suponhamos que os circuitos responsáveis pela percepção de cores do cérebro de Atílio sejam transplantados para o cérebro de Hipólito. Lembremo-nos de que os circuitos de Atílio são chips de silício e que os de Hipólito são constituídos por neurônios e sinapses. Contudo, a cirurgia é feita de modo tal que os circuitos de Hipólito são preservados. O cirurgião instala um interruptor na cabeça de Hipólito: quando esse é ligado, ativam-se os circuitos de silício e Hipólito percebe as cores como se fosse Atílio. Os circuitos de silício são cuidadosamente instalados na cabeça de Hipólito, de forma a evitar eventuais problemas. Note-se também que ambos os circuitos – o de neurônios e o de chips de silício – são funcionalmente equivalentes.

O que ocorre quando o circuito de silício é ligado? Antes de ele estar ligado, Hipólito estava experienciando algo vermelho. Quando o interruptor é ligado, Hipólito pas-

sa a perceber o mundo como Atílio, isto é, o que era experienciado como vermelho passa a ser verde. Ou seja, subitamente, o que era vermelho torna-se verde para Hipólito. Alguém poderia brincar com o interruptor, causando uma oscilação na percepção de Hipólito, que ora veria um objeto como sendo verde, ora o mesmo objeto como sendo vermelho. Como os circuitos de silício e os de neurônios são funcionalmente equivalentes, Hipólito *não deveria perceber* a mudança de verde para vermelho e vice-versa. Pelo menos seria isto o que deveria acontecer se o argumento de Shoemaker estivesse correto, pois a equivalência funcional não é suficiente para detectar ou determinar diferenças nas experiências subjetivas. Hipólito, quando recebe os circuitos de silício de Atílio e quando esses são ligados, transforma-se numa variante do robô de que nos fala Block: ele *deveria* ser incapaz de perceber essas diferenças de experiência subjetiva. Mas ao se concluir isto surge um paradoxo similar àquele que apontamos para o caso dos *absent qualia*: Hipólito *não poderia* perceber a variação de suas experiências de cor, a dança entre vermelho e verde quando seus circuitos de silício são ativados. A oscilação entre vermelho e verde não poderia ser percebida, pois esses correspondem a estados funcionalmente equivalentes entre circuitos de neurônios e circuitos de silício.

Mas essa é uma hipótese inaceitável: como ele não poderia perceber a variação de suas próprias experiências? Ele não poderia estar errado acerca de suas experiências; supor essa última possibilidade nos conduz, como no caso dos *absent qualia*, a um absurdo.

Mais algumas respostas

Examinemos agora que tipo de resposta um funcionalista poderia dar ao argumento do quarto do chinês. Esse também é um tópico que tem merecido muita atenção na filosofia da mente das últimas décadas. Há inúmeros contra-argumentos à objeção formulada por Searle[9], mas apresentaremos apenas um, inspirado no trabalho de Copeland (1993).

Copeland sugere que Searle comete um erro de raciocínio na sua formulação do argumento do quarto do chinês – um erro de inferência das partes em direção ao todo, também conhecido como "falácia das partes para o todo". Na montagem de seu experimento mental, Searle teria enfatizado excessivamente o papel da pessoa que está trancada no quarto e que recebe, sistematicamente, histórias e perguntas sobre essas histórias, todas elas em chinês. Searle se pergunta se o ocupante do quarto entende ou não chinês, mas nunca se *o sistema como um todo* entende chinês – o que romperia com a concepção antropomórfica de compreensão por ele pressuposta, alterando significativamente os resultados que poderiam ser derivados de seu experimento mental.

9. Outras objeções são apresentadas em Teixeira (1998).

Quando falamos no *sistema como um todo* estamos nos referindo ao quarto, com suas aberturas por onde entram e por onde saem mensagens escritas em chinês, além, é claro, do ocupante. Mas note-se que o ocupante é apenas uma parte desse sistema, embora seja, sem dúvida, uma parte importante. Por que então perguntar apenas para o ocupante se ele entende chinês e não indagar se o sistema como um todo poderia entender chinês? Questionar apenas se o ocupante do quarto entende ou não chinês, em vez de perguntar se o quarto + as aberturas + o ocupante (o sistema como um todo) entende chinês, equivaleria a perguntar se alguns neurônios (e quais?) de uma pessoa entendem chinês e não se essa pessoa entende ou não chinês. Ao ignorar isto, Searle comete a falácia de inferir da premissa de que o ocupante do quarto (uma parte do sistema) não entende chinês que *o sistema como um todo também não poderia entender chinês*.

Em outras palavras, Searle estaria fazendo o seguinte raciocínio:

Premissa: A manipulação de símbolos feita pelo ocupante do quarto não o capacita a entender chinês.

Conclusão: Logo, a manipulação de símbolos feita pelo ocupante do quarto não capacita tampouco o sistema como um todo a entender chinês.

Suponhamos que o ocupante do quarto se chame Jorge. Um equivalente do raciocínio de Searle seria então o seguinte:

Premissa: Jorge é vendedor da fábrica de cuecas da marca Belmonte.

Jorge nunca vendeu cuecas da Belmonte para a Guatemala.

Conclusão: A Belmonte nunca vendeu cuecas para a Guatemala.

Searle certamente tem uma réplica para esse tipo de ataque ao argumento do quarto do chinês. Ele sugere o seguinte: suponhamos que o ocupante do quarto, o Jorge, memorize todas as regras de conversão e faça todas as operações de associação simbólica na sua cabeça. Ele não precisaria então estar trancado no quarto, na verdade ele passaria a ser o sistema como um todo. Ora, se ele não entende chinês, o sistema como um todo tampouco entende chinês, pois esse último passaria então a ser apenas uma parte de Jorge.

Nesse caso, para se tornar válido, o argumento do quarto do chinês precisa incluir:

a) O sistema é parte de Jorge

b) Se Jorge não entende chinês então nenhuma parte de Jorge tampouco entende chinês.

Mas será essa réplica suficiente para manter a validade do argumento do quarto do chinês? Suponhamos que Jorge seja raptado por uma gangue de fanáticos da inteligência artificial e que essa opere seu cérebro, implantando numa parte de seu córtex um "programa neuronal". O "programa neuronal" é capaz de provar uma série de teoremas lógicos, e é implantado sem danificar nenhuma parte do cérebro de Jorge. De tempos em tempos, Jorge recebe uma série de inputs – proposições matemáticas e lógicas –

que ele não compreende, mas seu cérebro é capaz de produzir provas matemáticas para essas proposições, ou seja, uma série de outputs cuja natureza Jorge tampouco é capaz de compreender. Ele fica olhando, assustado, para esses outputs. Jorge não pode demonstrar teoremas da lógica – ele não sabe como fazer isto –, mas uma parte de seu cérebro executa essa tarefa de forma eficiente.

Isto equivale a uma situação do seguinte tipo: alguém pergunta para Jorge: "Você sabe demonstrar teoremas da lógica"? E Jorge responde: "não, eu não sei, mas alguns de meus neurônios sabem". É esse mesmo tipo de paradoxo que é construído por Searle ao centrar seu argumento do quarto do chinês na indagação de se o falante trancado no quarto entende ou não, em vez de considerar o sistema como um todo. Não tem cabimento perguntar se os neurônios de uma pessoa (parte de um sistema) entendem chinês, mas se a pessoa entende chinês ou não. A pergunta só adquire sentido quando consideramos que os neurônios são parte da pessoa, e é isto que pressupomos ao perguntar para uma pessoa se ela entende chinês, esperando que ela responda "sim" ou "não". Não dirigimos a questão para os neurônios ou para um grupo específico de neurônios dessa pessoa, mesmo que admitamos que a capacidade de entender chinês possa, em última análise, se dever à atividade específica de algumas partes de seu cérebro. É bem provável que a capacidade de entender chinês não se deva a partes específicas do cérebro de alguém e envolva a atividade simultânea de várias partes de seu cérebro, o que nos impediria, em última análise, de dizer que são alguns neurônios que entendem chinês. Se o pressuposto localizacionista é abandonado (isto é, a ideia de que funções específicas são desempenhadas por partes específicas do cérebro) a ênfase em partes específicas do sistema (o falante trancado no quarto) como critério para se saber se esse sistema entende ou não chinês deve também ser abandonado. E sem essa ênfase, o argumento de Searle não parece fazer muito sentido.

A teoria dos sistemas intencionais

A teoria dos sistemas intencionais constitui uma versão específica do funcionalismo proposta pelo filósofo norte-americano Daniel Dennett no início dos anos de 1980. Dessa nova versão do funcionalismo, Dennett deriva também um novo modo de conceber as relações entre mente e cérebro. Podemos entender sua proposta através de uma alegoria.

Suponhamos que um dia um objeto estranho, parecido com uma máquina, caísse do céu, como um meteorito. Por uma curiosidade natural, você, como bom cientista, procuraria examinar esse objeto, desmontando-o para ver como e com que material foi construído. O objeto parece ser uma máquina complexa, com diversas partes móveis, ligadas a uma caixa que – ao que tudo indica – seria uma espécie de controlador central. Os materiais utilizados para construir esse tipo de máquina, bem como a energia por ela utilizada não diferem muito daqueles que encontramos no planeta terra. A máquina liga e desliga com eletricidade, tem um *plug* igual a qualquer um dos nossos eletrodomésticos caseiros, contudo, a observação dos movimentos da máquina não nos permite saber *para que* serve esse tipo de mecanismo. Há apenas uma marca, uma espécie de

emblema gravado num canto da máquina, um tipo de etiqueta que você nunca viu antes.

Alguns meses depois uma nave espacial desce no seu quintal. A curiosidade faz com que você se aproxime dessa nave. Nesse momento você é sugado para dentro dela, numa típica situação de rapto interplanetário que vemos em filmes. Dentro da nave você verifica que há vários painéis nos quais estão gravados emblemas idênticos àquele que você viu na máquina que tinha caído do céu – a tal máquina que você não sabia para que servia; embora conhecesse sua composição química e todos os seus mecanismos. Obviamente você deduz: "estão me transportando para o mesmo planeta de onde veio aquela máquina".

Quando você desembarca nesse planeta desconhecido, a primeira coisa com a qual você se depara são diversas máquinas daquele tipo, operando em vários lugares. São máquinas que servem para aparar um tipo especial de flor, extremamente dura e que cresce em vários lugares desse planeta. A máquina servia para cortá-las e imediatamente triturá-las, pois, ao que tudo indicava, eram flores daninhas, com um caule extremamente rígido e difícil de ser cortado. Você então percebe: "Ah, a máquina era um cortador de flores". Mas será que você poderia ter deduzido isto antes de ver a máquina operando num tipo de meio ambiente completamente diferente? Será que você teria sido capaz de saber *para que* servia a máquina antes de vê-la operando?

O mais estranho nesse planeta era o fato de ele ser habitado por criaturas fisicamente parecidas com os seres humanos (vamos chamá-las, por convenção, de *critters*). O único problema – bastante grave, aliás – era de você não ter a menor noção de como se sucediam os comportamentos dessas criaturas. Em outras palavras, os comportamentos dessas criaturas eram *absolutamente ininteligíveis*. Após várias semanas de angústia, você – como bom cientista – decide desenvolver uma *ciência do comportamento* dos critters. Mas aí começam a aparecer vários problemas.

Em primeiro lugar, o comportamento deles era muito irregular, quase imprevisível (e isto certamente já criava também uma série de dificuldades práticas para se sobreviver naquele planeta). Em segundo lugar, mesmo quando alguns desses comportamentos se sucediam com regularidade acontecia o mesmo que ocorrera com a tal máquina: não dava para saber *para que* esses comportamentos serviam. E certamente os critters tinham uma cultura e civilização muito desenvolvidas, o que fazia supor que seus comportamentos fossem bastante complexos, ou seja, que eles não se resumiam simplesmente em comer e se reproduzir. Em terceiro lugar, os critters eram dotados de um cérebro extremamente complexo. Seus cérebros tinham, aliás, uma peculiaridade marcante: um mesmo comportamento podia ser produzido por diferentes tipos de eventos neuronais e até mesmo por diversos tipos de reações bioquímicas. Em outras palavras, não era possível estabelecer conexões regulares, uma a uma, entre os comportamentos dos critters e o que se passava nos seus cérebros.

Ora, como poderia ser desenvolvida uma ciência do comportamento dos critters? Ou, pelo menos, um tipo de conhecimento que tornasse seus comportamentos inteligíveis? (O que sem dúvida resolveria vários problemas práticos nas tentativas de intera-

ção com eles). Para desenvolver essa ciência do comportamento, por mais simples que seja, seria necessário *montar uma história* a partir da observação de seus comportamentos que, apesar de serem irregulares, deveriam se prestar à detecção de alguns padrões. Ora, como seria possível estipular esses padrões, se os seus comportamentos envolviam uma grande margem de imprevisibilidade?

Montar essa história requer uma *reconstrução racional* dessas sequências de comportamentos. Essa reconstrução racional só se torna possível se nossa história incluir um conjunto de *conceitos articuladores* que ordenem esses comportamentos apesar de sua imprevisibilidade. Esses conceitos articuladores podem ser obtidos a partir do momento em que atribuamos aos critters *intenções*, *crenças*, *desejos*, etc. A partir daí começamos a construir uma verdadeira *psicologia* critteriana. Usando as noções de crença, desejo, intenção etc. e como essas se inter-relacionam numa espécie de sistema hipotético aumentamos nossa capacidade de predizer alguns de seus comportamentos. Ou seja, diminuímos sua imprevisibilidade à medida que passamos a compreendê-los melhor. Ao atribuir intenções a seus comportamentos livramo-nos também do mesmo tipo de perplexidade que nos atormentou quando encontramos no nosso quintal máquinas desconhecidas que caíram do céu. Máquinas que não sabemos para que foram construídas, mesmo depois de desmontá-las e examinar seus mecanismos e os materiais que as compõem.

Depois de alguns ajustes iniciais o manual improvisado de psicologia critteriana começa a funcionar bem. Aliás, funciona tão bem que depois de alguns meses ocorre um fenômeno peculiar: você começa a acreditar que os critters realmente têm intenções, crenças, desejos etc. Como bom cientista, você então vai tentar um novo empreendimento: examinar o cérebro dos critters para verificar onde estão localizadas suas intenções, crenças etc. Inicialmente, tudo parece correr bem, pois o cérebro dos critters é quase igual ao nosso. Mas aí você se depara com um grande problema: por mais que você examine seu tecido neuronal, não encontra nada parecido com intenções, crenças etc. Tudo o que você vê são neurônios, conexões entre esses, correntes elétricas e várias reações químicas. De nada adiantou você estudar o funcionamento e a composição química do cérebro dos critters.

A perplexidade torna-se ainda maior quando você percebe que, por ter começado a realmente acreditar que os critters têm intenções, crenças, desejos etc., ou seja, todo um *vocabulário de termos mentais* você acaba de gerar um problema filosófico: a versão critteriana do problema mente-cérebro. Obviamente esse é um problema insolúvel: como as intenções, crenças, desejos etc. foram uma criação sua (embora você tenha se esquecido disto após alguns meses) nunca será possível estabelecer um mapeamento entre a mente-critter e o cérebro-critter. Nem entre a psicologia-critter e a neurociência-critter.

Mas nossa história não termina aí. Depois de um certo tempo você faz uma descoberta ainda mais surpreendente: os critters mantêm, dentro de sua universidade, um Centro de Ciências Humanas. Talvez o único que mereça realmente ter esse nome, pois de lá os critters observam e estudam os humanos da mesma maneira que nós ob-

servamos e estudamos ratos e pombos nos laboratórios de psicologia experimental. Ao visitar a biblioteca do Centro de Ciências Humanas você descobre que lá há um único livro – um volume imenso. Na capa está escrito: "Manual completo da ciência do comportamento humano". Afoito, você abre a primeira página, onde vem escrito: "Esse livro nunca foi e nunca será traduzido para nenhum idioma humano. Trata-se de um texto *intraduzível* para idiomas humanos". Indignado e frustrado, você vai falar com o diretor do Centro de Ciências Humanas. O diretor diz para você: "Não adiantaria nada tentar traduzir esse manual para um idioma humano, embora nós saibamos tudo sobre o cérebro humano, sobre o comportamento de seres humanos e como o vosso cérebro produz aquilo que vocês chamam de mente etc. O problema é que o manual *não faria sentido* para vocês, pois, para entendê-lo, vocês teriam de deixar de ser humanos para se observarem da mesma maneira que vocês observam outras espécies no seu planeta. Mas isto certamente é impossível".

Ora, a esta altura alguém poderia, com toda razão, perguntar: mas a que vem toda essa história, cheia de desencontros, paradoxos e até mesmo de contradições? O que isto tem a ver com o funcionalismo de Dennett? A ideia central do funcionalismo de Dennett consiste em sustentar que nossos estados mentais, sobretudo as intenções, crenças, desejos etc. (os elementos que compõem a chamada *folk psychology* ou psicologia popular e a partir dos quais construímos nossas explicações habituais dos comportamentos dos outros seres humanos), nada mais são do que um sistema hipotético de conceitos articuladores que utilizamos para tornar inteligíveis os comportamentos de outros seres humanos. Esses conceitos articuladores desempenham na *folk psychology* o mesmo papel que os chamados *termos teóricos* das diversas teorias científicas.

Para entender o que são termos teóricos, basta pensar, por exemplo, na física. A física se utiliza de vários termos teóricos, como "massa" ou "centro de gravidade". Não podemos *observar* a massa de um corpo (embora possamos medi-la). Podemos determinar o centro de gravidade de um corpo, sistema etc., mas a própria noção de centro de gravidade é hipotética. Ela é uma construção teórica. Não há nenhum *objeto específico na natureza* que corresponda a um centro de gravidade. Sistemas físicos diferentes têm centros de gravidade diferentes e não faria sentido dizer que um centro de gravidade tem uma realidade própria, independente do sistema físico que estamos considerando. Contudo, essas duas noções – massa e centro de gravidade – têm um papel articulador fundamental que nos permite construir teorias físicas e a partir dessas fazer predições acerca do movimento dos corpos. Em outras palavras, "massa" e "centro de gravidade" são *ficções teóricas úteis*.

Ora, o mesmo ocorreria com intenções, crenças etc.; elas são, no entender de Dennett, termos teóricos ou ficções úteis que nos permitem explicar e predizer comportamentos de organismos e sistemas. Esses termos teóricos ou ficções úteis são chamados por Dennett de *posits* ou *abstracta*. *Posits* e *abstracta* estão na nossa mente e na nossa linguagem, formando um *sistema intencional*. Os elementos dos sistemas intencionais não têm correspondentes na natureza nem tampouco nos cérebros. Não há nada no

cérebro que corresponda a intenções, crenças etc., ou seja, não há nenhum estado cerebral específico que seja intrinsecamente dotado de sentido ou de significado. Sentido e significado são atribuições que fazemos a organismos ou sistemas na medida em que esses possam ser descritos como sistemas intencionais. Sentido e significado não são gerados pela atividade cerebral, mas resultam da descrição que fazemos dessa e de suas consequências, ou seja, nossos próprios comportamentos e os comportamentos dos organismos e sistemas que nos cercam. Quando esses comportamentos obedecem a padrões mínimos de racionalidade, passam a poder ser descritos como sistemas intencionais – sistemas aos quais atribuímos intenções, crenças e desejos para que seus comportamentos se tornem inteligíveis para nós. A possibilidade e a atividade pela qual montamos histórias inteligíveis do comportamento desses organismos e sistemas articulando-as a partir de termos teóricos como intenções, crenças, desejos etc. (a *folk psychology*) Dennett chama de atitude intencional (*intentional stance*).

Um aspecto interessante da teoria dos sistemas intencionais é o fato de que podemos atribuir intenções ou crenças (ou seja, atribuí-los a partir da atitude intencional) não apenas a seres humanos e outros organismos, mas também a sistemas artificiais desde que esses sejam capazes de produzir comportamentos dotados de padrões mínimos de racionalidade. A detecção de um padrão mínimo de racionalidade no comportamento de um organismo ou sistema é indício de que estamos diante de um sistema intencional, ou seja, um sistema que requer elementos da *folk psychology* para que possamos compreender seus comportamentos. É nesse sentido que o funcionalismo de Dennett tem exercido um grande fascínio sobre os teóricos da inteligência artificial. Pense, por exemplo, num jogador de xadrez artificial como o Deep Blue. Se o Deep Blue produzir comportamentos que se assemelhem aos de um jogador de xadrez humano (parece que o Deep Blue conseguiu produzir mais do que isto...) ele poderá ser descrito a partir da atitude intencional – pouco importando se no seu interior encontremos chips ou neurônios, pois a adoção da atitude intencional é perfeitamente compatível com a tese da múltipla instanciação (*multiple realizability*).

Ora, será então que, para Dennett, aquilo que chamamos de "mente" não passaria de uma ilusão que surge dos truques que utilizamos para descrever o comportamento de outros organismos ou sistemas artificiais quando não conseguimos explicar e prever inteiramente seus comportamentos? Essa pergunta precisa ser respondida em duas etapas.

Em primeiro lugar, é preciso dizer que Dennett não considera que os termos teóricos da nossa *folk psychology* sejam meras ilusões. Num de seus mais importantes artigos, "Real Patterns" [Padrões reais], publicado em 1991 (1991a), ele chama a atenção para o fato de que precisamos distinguir entre *meras ficções* e *ficções úteis*. Nem todas as ficções são úteis, muito menos necessárias. Ficções tornam-se úteis e importantes na medida em que elas aumentam o grau de previsibilidade das teorias científicas, sejam elas teorias físicas ou teorias psicológicas – como é o caso da nossa *folk psychology*. Pensemos, por exemplo, no caso do princípio de inércia, uma das mais importan-

tes ficções teóricas da mecânica clássica. O princípio de inércia diz que um corpo permanecerá em repouso se nenhuma força for aplicada sobre ele. Contudo, se uma força for aplicada sobre esse corpo, esse permanecerá em movimento indefinidamente. Suponhamos que esse corpo seja uma bola. Será que alguém já viu uma bola que, ao receber um impulso, permanece em movimento *indefinidamente*? É óbvio que não. Contudo, sem essa ficção jamais poderíamos ter formulado uma teoria tão importante quanto a mecânica clássica!

Mente, cérebro e algoritmos de compressão

Dennett distingue três níveis de explicação quando tentamos entender o comportamento de um organismo ou sistema: o nível físico (*physical stance*), o nível do *design* (*design stance*) e o nível intencional (*intentional stance*). Pensemos agora num computador capaz de jogar xadrez. O nível físico corresponde à descrição dos materiais que compõem cada uma de suas partes, até se chegar, por exemplo, às partículas subatômicas que se encontram nas suas peças de silício. O nível do *design* corresponde à descrição da arquitetura desse computador e como as peças se ligam umas às outras permitindo que ele funcione dessa ou daquela maneira. O terceiro nível – o intencional – corresponde à descrição que fazemos do "comportamento" desse computador quando está jogando xadrez. Se ele for um bom jogador de xadrez, identificaremos, a partir de suas jogadas, algum tipo de racionalidade e, a partir dessa, começaremos a atribuir a essa máquina algum tipo de predicado mental como, por exemplo, "ser inteligente" ou "inteligência". O nível intencional surge à medida que podemos montar uma história inteligível acerca dos comportamentos dessa máquina, e, para montá-la, precisamos ir progressivamente utilizando conceitos articuladores como "racionalidade" ou "inteligência". Nossos computadores atuais ainda estão longe de serem máquinas para as quais atribuiríamos predicados mentais, mas nada nos impede que no futuro possamos nos aproximar disto.

Esses três níveis de explicação são compatíveis, mas, ao mesmo tempo, irredutíveis entre si. Não poderíamos entender o funcionamento de um motor a combustão (*design stance*) estudando as moléculas que compõem seus cilindros e pistões (*physical stance*). Da mesma maneira, é pouco provável que possamos compreender o funcionamento do cérebro humano descrevendo-o unicamente sob o aspecto molecular. Mas será o nível intencional redutível ao nível do *design* e ao nível físico? Em outras palavras, serão os conceitos articuladores da *folk psychology* que compõem a atitude intencional redutíveis a algum tipo de estrutura cerebral?

À primeira vista poderíamos responder positivamente a essa questão. A adoção da atitude intencional pode ser vista como a expressão de nossa ignorância e da consequente incapacidade de prevermos o comportamento de um organismo ou sistema. Se conhecêssemos inteiramente o funcionamento desse organismo ou sistema poderíamos tornar seu comportamento inteligível e efetuar predições acerca de seus desenvolvimentos futuros a partir do nível do *design* e do nível físico. Esse tipo de passa-

gem, de uma descrição sob o aspecto intencional para uma descrição sob o aspecto do *design* ou sob o aspecto físico já teria ocorrido na história da humanidade: trocamos o animismo primitivo que atribuía intenções e desejos à natureza e à atmosfera pela explicação científica do clima que nos é proporcionada pela meteorologia e pela física. Poderíamos supor que o mesmo ocorrerá, algum dia, no que diz respeito à explicação do comportamento humano à medida que as neurociências forem se desenvolvendo.

Por outro lado, podemos nos indagar se seríamos capazes, algum dia, de conseguir uma descrição física tão completa de nós mesmos a ponto de podermos prescindir de qualquer descrição intencional. Ora, isto não equivaleria ao sonho do "Manual completo da ciência do comportamento humano" de autoria dos cientistas critters? Será que nós, humanos, não precisaríamos transcendermos a nós mesmos para poder escrever esse manual – tornar-nos seres de outros planetas para observarmos seres humanos como objetos de estudo de uma ciência do comportamento? E ao fazer isto não correríamos o risco de tornar essa ciência ininteligível para nós mesmos?

Mas mesmo que esse tão sonhado manual possa um dia ser escrito em linguagem inteligível, ainda assim é possível que ele não dê conta do recado. É possível que a descrição do funcionamento de nosso cérebro e sua interação com o meio ambiente (incluindo outros seres humanos) não nos permita efetuar predições corretas, mas apenas aproximadas, acerca do comportamento humano, da mesma maneira que a meteorologia só pode fornecer predições do tempo dentro de uma margem de probabilidades. Da mesma maneira que o clima, o comportamento humano forma, ao que tudo indica, um sistema complexo que oscila entre regularidades e modificações abruptas – aquilo que os físicos chamam de sistemas dinâmicos complexos. Nesse caso, a *folk psychology* continuaria sendo necessária para entendermos o comportamento de outros seres humanos: a despeito de estarmos de posse de nosso manual completo de psicologia humana, ainda assim precisaríamos recorrer à atribuição de crenças, desejos, intenções etc. para podermos tornar essas irregularidades inteligíveis.

Contudo, há uma outra possibilidade que não pode ser descartada: a de que nosso manual seja tão sofisticado que se consiga, a partir dele, previsões com uma margem quase total de acerto. Além disto, esse manual incluiria um mapeamento integral entre os termos da *folk psychology* e as suas estruturas cerebrais correspondentes. Seria isto o fim da *folk psychology* pela sua redução ao nível do *design* e ao nível físico? Ora, mesmo que isto ocorra, não significa que a *folk psychology* deva necessariamente desaparecer: o mais provável é que ela continue a existir, *apesar dessas reduções*. O mapeamento integral do cérebro, incluindo a própria *folk psychology* nos descortinaria um conhecimento tão complexo que tornaria nossa vida cotidiana impraticável. Imagine o que seria nossa vida se para tomarmos uma decisão simples, do tipo "o que devo fazer em seguida", eu tivesse de percorrer todas as redes ativas que compõem o pensamento, neurônio por neurônio e, em seguida, molécula por molé-

cula, todos eles bombardeados a cada instante por estímulos externos produzindo mais e mais episódios microscópicos... a *folk psychology* nos proporciona um atalho muito mais curto! É provável – como sugerimos no capítulo anterior – que a complexidade do nosso cérebro seja um horizonte insuperável, e que, nesse caso, estejamos confinados à *folk psychology*.

Dennett supõe que a *folk psychology* seja uma das mais importantes conquistas evolucionárias da espécie humana – uma conquista que nos permite lidar com o meio ambiente e com outros seres humanos apesar de sua complexidade insuperável. A *folk psychology* seria uma coleção natural de ficções úteis que permite organizar nosso comportamento e o comportamento de outros organismos e sistemas de modo a diminuir sua margem de imprevisibilidade. Mas como nossas intenções, crenças e desejos poderiam existir como tal e, ao mesmo tempo, corresponderem a estruturas cerebrais? Como termos teóricos, ou seja, *posits* ou *abstracta* [abstrações] podem ter uma existência própria e ao mesmo tempo ser algo no nosso cérebro?

Nossa tendência habitual seria supor que os elementos da *folk psychology*, por serem apenas instrumentos, não poderiam ter qualquer realidade própria; seriam meras "construções mentais" ou meras "interpretações" atribuídas aos comportamentos de outros organismos e sistemas. Postular sua existência independentemente de estruturas cerebrais equivaleria a romper com o materialismo. Por outro lado, postular sua redutibilidade a estruturas cerebrais equivaleria a eliminá-las num futuro próximo – seria o mesmo que apostar no materialismo eliminativo. Nenhuma dessas alternativas é seguida por Dennett, e nisto reside seu tratamento original do problema das relações entre mente e cérebro. Há uma esfera própria de descrição psicológica dos organismos, e essa esfera corresponde à atitude intencional. Poderemos, no futuro, encontrar os correlatos neurais dos elementos da *folk psychology*, mas isto nos revelaria muito pouco acerca de sua natureza e de sua função.

É bem pouco provável que possamos *reduzir* os elementos da *folk psychology* a estruturas cerebrais como querem os partidários do materialismo reducionista, mesmo que utilizemos técnicas avançadas de mapeamento cerebral como aquelas proporcionadas pela neuroimagem (abordaremos essas técnicas no capítulo seguinte). Em outras palavras, em nenhum momento podemos prescindir da interpretação intencional ou da atitude intencional; tudo se passa como se estivéssemos irremediavelmente confinados a elas. O conhecimento de nossa mente é limitado pelo tipo de descrição que podemos fazer dela; nossa cognição não pode superar a si própria a ponto de podermos ler e entender o "Manual completo da ciência do comportamento humano" que algum critter ou marciano teria escrito.

Que tipo de realidade têm, então, os elementos que compõem a *folk psychology*? Intenções, crenças, desejos etc., são interpretações, mas essas não estão apenas na mente ou na cabeça daqueles que observam os comportamentos de organismos ou sistemas. O que lhes confere realidade – uma realidade a meio caminho entre a pura e simples construção mental ou subjetiva e a sua existência como estrutura cerebral – é o fato de essas interpretações captarem e expressarem *padrões* ou *regularidades* que es-

tão na natureza. Os comportamentos dos organismos, embora não inteiramente previsíveis, oscilam entre mudanças abruptas e regularidades e é a partir da detecção destas últimas que tentamos torná-los menos imprevisíveis. É esse aspecto da *folk psychology* – sua correspondência a esses padrões – que a torna um poderoso instrumento de sobrevivência.

A atribuição de intenções, crenças e desejos torna-se, assim, um instrumento e uma estratégia a partir da qual podemos contornar a extraordinária complexidade presente no cérebro e no comportamento de outros organismos. Nesse sentido, os elementos da *folk psychology* funcionam como verdadeiros "algoritmos de compressão", a partir dos quais podemos apreender rapidamente os padrões ou regularidades do comportamento. Para termos uma ideia do que seja um "algoritmo de compressão" basta que imaginemos uma situação na qual queiramos "escanear" uma figura para, em seguida, tentar salvá-la num disquete de 1,44 MB. Se a figura for muito grande, ela não caberá no disquete e, a não ser que tenhamos um "zip-drive", não poderemos transportá-la para um outro computador para, por exemplo, inseri-la num texto disponível nesse último. Contudo, aqueles que têm experiência nesse tipo de situação poderão lançar mão de um artifício: compactar a figura para poder transportá-la em disquete, para, em seguida, descompactá-la no computador de destino. Para compactar a figura usa-se um "algoritmo de compressão" – ele "diminui" a figura temporariamente, identificando nessa padrões ou regularidades que podem ser "comprimidos" ou "compactados".

O uso de "algoritmos de compressão" constitui uma estratégia computacional para contornar a complexidade que está presente na figura e que impede que essa seja salva no espaço reduzido que temos no disquete – compactar a figura é um primeiro passo para contornar sua complexidade *quantitativa*. Note-se que somente padrões ou repetições podem ser comprimidos; se esses não ocorrerem na figura (embora isto seja difícil de imaginar) sua versão compactada será idêntica à original.

Da mesma maneira que os algoritmos de compressão, os elementos da *folk psychology* permitem contornar a complexidade e imprevisibilidade do comportamento de outros organismos e sistemas, compactando padrões e regularidades presentes no seu cérebro e no seu comportamento, ou seja, possibilitando sua apreensão e descrição em tempo real. E, da mesma maneira que nas tentativas de comprimir figuras, haverá elementos que não podem ser apreendidos e compactados por interromperem ou não corresponderem a nenhum tipo de regularidade ou padrão. A construção de uma história inteligível do comportamento de um organismo complexo – usando os recursos da *folk psychology* – será, igualmente, pontuada por uma oscilação entre regularidades e pontos de inflexão que caracterizam os algoritmos de compressão.

O QUE LER

Sobre funcionalismo:

BLOCK, N. *Readings in the Philosophy of Psychology*, vol. I

Verbetes sobre "Funcionalismo" em GUTTENPLAN, S. (org.) *A Companion to the Philosophy of Mind*

Sobre críticas ao funcionalismo:

SEARLE, J. *Mind, Brain and Science*

Sobre o funcionalismo de D. Dennett:

DENNETT, D. *Content and Consciousness*

DENNETT, D. *Brainstorms* (especialmente o capítulo I)

CAPÍTULO VII

TEORIAS DA CONSCIÊNCIA

O leitor atento deve ter notado que até agora estivemos usando as palavras "mente", "estados mentais" e "experiência consciente" como designando a mesma coisa. Em nenhum momento marcamos uma diferença entre "mente" e "consciência". Seguindo essa tendência histórica da filosofia da mente do século XX, não enveredamos por essa discussão. Até a metade dos anos de 1980, estabelecer essa distinção não parecia ser uma preocupação visível entre os filósofos da mente. Havia grande entusiasmo com as perspectivas abertas pela inteligência artificial e com a possibilidade de simulação mecânica das atividades mentais humanas através da construção de mentes artificiais. Pouco importava se uma máquina de jogar xadrez sabia ou não o que estava fazendo. O tema "consciência" não fazia parte da agenda dos funcionalistas, pois, para esses, processamento de informação e experiência consciente eram dissociáveis. Mas seria possível simular a cognição humana sem simular, ao mesmo tempo, seu aspecto *consciente*? Não seria essa uma diferença essencial entre mentes artificiais e humanas?

Essas foram as perguntas que começaram a ser formuladas no final da década de 1980. Tudo se passava como se a simulação da atividade mental humana fosse uma tarefa perfeitamente exequível, dependendo apenas de avanços tecnológicos. Restaria apenas saber o que tornaria um estado mental algo consciente, e, para isso, seria necessário responder algumas questões que não deixavam de causar perplexidade: "O que é consciência?", "Que tipo de papel desempenha essa na explicação da cognição humana?" "Existe cognição sem consciência?" "Terá a consciência um papel causal na produção da cognição e do comportamento?" "Podemos tratar a questão da consciência como um problema científico, isto é, como um problema empírico?"

Desde então, uma profusão de teorias acerca da natureza da consciência começou a proliferar na filosofia da mente. A tarefa a ser enfrentada era (e continua sendo) nota-

damente árdua: não existe nada mais imediato do que a experiência consciente, mas, ao mesmo tempo, não existe nada tão difícil de ser explicado.

As tentativas de montar estratégias para responder a essas perguntas produziram uma divisão de águas entre grupos de filósofos que professam opiniões conflitantes. De um lado estão os naturalistas, aqueles que acreditam poder explicar a natureza da consciência através de teorias computacionais ou através do estudo do funcionamento cerebral. Filósofos e cientistas cognitivos como R. Jackendoff (1987), os Churchlands (1986, 1995), W. Calvin (1990), D. Dennett (1991), O. Flanagan (1992) seguem essa tendência, apostando no triunfo de teorias materialistas e aliando-as, às vezes, ao darwinismo. Os não naturalistas como Swinburne (1984) e Nagel (1974, 1986) adotam uma posição radicalmente oposta: para eles *qualia* e experiências conscientes são intratáveis do ponto de vista de qualquer tipo de teoria neurocientífica. A impossibilidade de formular uma teoria científica da consciência torna o problema mente-cérebro insolúvel e força uma ruptura com o monismo materialista.

Formou-se ainda um terceiro grupo: os chamados "novos misterianos" (*new mysterians*). Esses não descartam a hipótese naturalista, mas sustentam que desvendar a natureza da consciência constitui um problema cuja complexidade ultrapassa a capacidade cognitiva humana. Os seres humanos podem até formular o problema da consciência – o que os distinguiria de outros animais –, mas seu cérebro é incapaz de resolvê-lo. Dentre os novos misterianos destaca-se McGinn (1991). McGinn sustenta que o cérebro humano, forjado pela evolução para sobreviver numa sociedade pré-industrial, não pode resolver esse tipo de questão, da mesma maneira que insetos não podem compreender a teoria da relatividade ou resolver uma equação diferencial. Não podemos transpor os limites de nossa própria razão. Na filosofia da mente, da mesma maneira que na matemática, há problemas cuja solução é impossível.

A predominância crescente do naturalismo gerou uma espécie de curto-circuito na história da filosofia da mente nas últimas décadas: de teorias da mente saltou-se para teorias da consciência, para retornar rapidamente para uma equiparação entre consciência e mente, num movimento quase imperceptível. Esse movimento se inicia no final dos anos de 1980 quando se enfatizou a necessidade de elaborar uma teoria da consciência por acreditar-se que uma teoria da mente não seria suficiente para explicar a natureza da experiência consciente. Uma teoria filosófica da natureza da experiência consciente tinha de começar por definir e caracterizar esse tipo de fenômeno, o que significava uma dificuldade quase intransponível. Para começar, seria preciso, pelo menos, unificar a pluralidade de sentidos que se pode atribuir ao termo "consciência" o que, aparentemente, não foi possível. Os filósofos da mente foram acusados de ficar girando em círculos, num exercício especulativo árido e inútil, onde nunca se chegou a qualquer tipo de consenso que servisse de ponto de partida para a elaboração de algum tipo de teoria.

O naturalismo rejeitou essa estratégia filosófica baseada na análise conceitual, acentuando que, em vez de ficarmos perguntando *o que é* a consciência – a busca por

uma resposta para o *hard problem* de que nos fala Chalmers[1] – temos de tratar essa questão como um problema científico, isto é, como um *problema empírico*. Em vez de definir consciência, precisamos estudar suas manifestações. Os Churchlands, por exemplo, propunham uma estratégia do tipo "dividir para dominar": em vez de tentarmos elaborar uma teoria geral da consciência, temos de elaborar teorias específicas de processos mentais nos quais essa se manifesta, ou seja, teorias acerca da natureza da atenção, da memória, dos processos cerebrais subjacentes à produção do sono e da vigília etc. Quando desvendássemos todos esses aspectos da nossa vida mental, encontrando seus correlatos neurais estaríamos de posse de uma teoria da consciência.

Ora, esse tipo de estratégia levou a uma equiparação entre teorias da consciência e teorias da mente num movimento inverso àquele que se iniciara no final dos anos de 1980. Uma teoria completa da mente seria, também, uma teoria da consciência. O resultado dessa manobra foi uma dissolução das fronteiras entre filosofia da mente e neurociência, que assistimos até hoje. Proliferaram teorias acerca dos correlatos neurais da consciência, tornando a tarefa de discorrer sobre todas elas quase impraticável. A filosofia da mente passou também a ser invadida por teorias oriundas da física, onde se busca uma equiparação entre experiência consciente e fenômenos quânticos, numa tentativa frequentemente ironizada como a busca da explicação do obscuro pelo mais obscuro.

Um quadro, ainda que incompleto, de todas essas teorias, remete-nos a uma lista razoavelmente longa dos possíveis correlatos neurais da consciência[2]:

– oscilações de 40 hertz no córtex cerebral (CRICK & KOCH, 1990).

– Núcleo intralaminar no tálamo (BOGEN, 1995).

– Mapas reentrantes nos sistemas tálamo-corticais (EDELMAN, 1989).

– Atividade neural no tempo (LIBET, 1993).

– Alguns neurônios no sulco temporal superior (LOGOTHETIS & SCHALL, 1989).

– Atividade rítmica em 40 hertz nos sistemas tálamo-corticais (LLINAS, RIBARY, JOLIOT & WANG 1994).

– *Gestalts* neuronais num epicentro (GREENFIELD, 1995).

– Coerência quântica nos microtúbulos do cérebro (HAMEROFF, 1994).

De todas essas hipóteses, as que se tornaram mais populares foram as de Edelman e as de Crick e Koch. Crick, ganhador de um Prêmio Nobel, popularizou-a no seu livro *The Astonishing Hypothesis* [A hipótese assombrosa], publicado em 1994. Nesse livro Crick chamou de hipótese assombrosa a possibilidade de explicar a natureza de nossos

1. Introduzimos essa noção no capítulo IV.

2. Para uma lista completa cf. Chalmers (1996b).

pensamentos, alegrias, tristezas e outras emoções como resultando da atividade de alguns grupos de neurônios de nosso cérebro. Não muito tempo depois, várias resenhas e comentários desse livro apontaram para o fato de que, afinal de contas, ninguém mais considerava que essa fosse uma hipótese assim tão assombrosa[3].

Crick supôs que a chave para desvendar o mistério da consciência estaria no estudo dos mecanismos neurais subjacentes à organização da percepção visual. A hipótese que ele desenvolveu baseou-se no fato de que a consciência visual está correlacionada com uma oscilação, em 40 Hz, das camadas cinco e seis do córtex visual primário. Ou seja, quando o córtex visual reage à estimulação, alguns grupos de neurônios disparam de forma sincronizada.

Ora, seriam esses, efetivamente, os correlatos neurais da consciência? Certamente não poderíamos esperar que qualquer dispositivo físico com partes oscilando em 40 Hz (um rádio, por exemplo) torne-se, necessariamente, consciente. A oscilação sincronizada parece relacionar-se com a produção de experiências conscientes unicamente quando ocorre no cérebro. Ou seja, a oscilação sincronizada não poderia ser isolada ou dissociada de outros elementos presentes no cérebro, o que a torna necessária, mas não suficiente para explicar a produção de experiências conscientes. Esse é o chamado *linking problem* que ainda constitui um grande desafio para os neurocientistas.

A teoria de Edelman (1987, 1989, 1992) não se popularizou tanto como a de Crick e Koch, embora tenha atraído a atenção de neurocientistas e cientistas cognitivos. Para a formulação de sua teoria, o *darwinismo neural*, Edelman partiu de cinco ideias básicas acerca do funcionamento cerebral[4]. A primeira delas é que seria impossível para o genoma humano especificar inteiramente a estrutura do cérebro. As conexões sinápticas não são preestabelecidas e desenvolvem-se a partir da interação do cérebro com o seu meio ambiente e com processos bioquímicos endógenos. Em segundo lugar, os cérebros dos indivíduos apresentam diferenças em termos de estrutura e de conectividade. Como consequência, não há um mapeamento fixo entre tipos de estados mentais e tipos de estruturas neuronais. Em terceiro lugar, da mesma maneira que pressões ambientais selecionam membros mais aptos numa espécie, as informações que entram no cérebro selecionam grupos de neurônios reforçando as conexões entre eles. Esses grupos competem entre si na tentativa de criar representações eficazes, ou mapas, da grande variedade de estímulos que chegam do meio ambiente. Grupos que formam mapas bem-sucedidos predominam, ao passo que os outros definham. Em quarto lugar, grupos de neurônios podem desempenhar múltiplos papéis. Detetores de vermelho são ativados quando coisas vermelhas estão na minha frente. Contudo, eles podem também ser ativados para reconhecer rosa ou púrpura. Em quinto e último lugar, a perda de neurônios não implica, necessariamente, na perda de capacidade funcional do cérebro, salvo quando essa perda é massiva como no caso, por exemplo, da doença de Alzheimer.

---

3. Esse mesmo tipo de observação é feito, de forma jocosa, por Horgan (1996).

4. Para essa exposição, cf. Flanagan (1991) e Edelman & Tononi (1995).

Dadas essas características do cérebro, a consciência surge a partir da possibilidade do cérebro estabelecer uma distinção entre experiência interna e externa, ou seja, distinguir entre mudanças decorrentes de variações orgânicas internas ao organismo (sinais que acusam modificações endócrinas) e mudanças ocasionadas por sua interação com o meio ambiente. Forma-se, com isto, uma distinção preliminar entre interno e externo, entre "eu" e "não eu", a partir de uma segregação entre dois tipos de sistema nervoso efetuada pela seleção neuronal, um para efetuar o registro interno e outro para efetuar o externo. Distinguir entre o interno e o externo é tarefa vital para a sobrevivência de alguns organismos, mas isto não seria possível se essa segregação inicial não viesse acompanhada, posteriormente, de uma coordenação entre esses dois tipos de sistema nervoso.

Dessa coordenação teriam surgido as primeiras formas de percepção consciente. Contudo, mais um passo teria sido necessário para o surgimento da consciência entendida como um fluxo serial. Para isto foi necessária a formação de um sistema de memória capaz de projetar para o futuro experiências bem-sucedidas no passado e atualizar-se constantemente com base nos novos resultados obtidos. Esse sistema estaria implementado no cérebro na forma de "mapas reentrantes". A reentrada é uma espécie de realimentação entre mapas ou grupos de neurônios selecionados. Assim sendo, se um mapa A envia um sinal a um mapa B e esse responde com um segundo sinal, esse reentra em A. A reentrada permitiria a categorização, o aprendizado, a projeção para o futuro e a formação de conceitos, passos necessários para a passagem de uma consciência perceptiva primária para formas mais sofisticadas de experiência consciente, incluindo a própria autoconsciência.

As teorias de Crick e Koch e a de Edelman são consideradas marcos importantes nas tentativas de elaboração de uma abordagem naturalista da consciência. Contudo, nelas não encontramos uma explicação da natureza da experiência consciente, ou seja, elas não explicam o que, em última análise, torna um estado mental algo consciente. Crick e Koch identificam a experiência consciente com a organização da percepção, buscando uma explicação de sua unidade em mecanismos neurais subjacentes. No caso de Edelman, encontramos uma identificação implícita entre consciência e atenção. Ambas são teorias neurológicas da *mente* e não da consciência. Não se poderia esperar de uma abordagem naturalista uma explicação *do que é* a consciência (o *hard problem* de Chalmers) – essa não é a proposta de Edelman, nem de Crick e Koch. Mas poderíamos – ou *deveríamos* – esperar uma explicação de como e por que a consciência afeta a cognição, ou seja, que diferença faz ter experiências conscientes.

Flanagan (1998) enfatiza que uma teoria da cognição que não leve em conta o papel que a consciência desempenha na nossa vida mental será necessariamente incompleta. O papel da experiência consciente na cognição deve ser o ponto de partida de uma teoria da consciência. Reconhecer que a experiência consciente faz diferença no processamento de informação constitui, aliás, o primeiro passo a ser dado para se construir essa teoria – um passo que nos força a restabelecer uma distinção entre mente e consciência como ressaltaram os filósofos no final dos anos de 1980. O grande desa-

fio a ser enfrentado por uma teoria da consciência é escapar, de um lado, da especulação filosófica estéril e, de outro, de achar que uma teoria da mente seria automaticamente uma teoria da consciência – como querem os naturalistas influenciados pelo funcionalismo.

Podemos perceber que o papel da consciência no processamento de informação não é apenas um efeito colateral se compararmos jogadores de xadrez mecânicos e humanos, como é o caso de Deep Blue e G. Kasparov. Deep Blue venceu Kasparov em 1997 após uma longa disputa. Reconhecemos em Deep Blue uma máquina inteligente, apesar de ela ser totalmente inconsciente, o que reforçaria o pressuposto funcionalista de que inteligência e consciência podem ser dissociadas e que essa última seria apenas um efeito colateral dispensável. Mas dificilmente reconheceríamos em Deep Blue um modelo de como os seres humanos jogam xadrez ou processam informações de outros tipos. Como ele não tem experiências conscientes, não podemos sequer *imaginar o que é ser como o Deep Blue* – o que nos coloca numa situação muito mais radical do que aquela concebida por Nagel quando esse nos diz que não podemos saber o que é ser como um morcego. Em Deep Blue não reconhecemos nada parecido com a cognição humana, apesar de ele ter sido construído por uma equipe de engenheiros e programadores e, do fato de seu programa registrar milhares de jogadas e soluções para problemas de xadrez executadas por seres humanos nas últimas décadas. A inconsciência de Deep Blue torna sua "psicologia" totalmente opaca para nós.

Consideremos agora o jogador de xadrez humano, ou seja, Kasparov. Sabemos o quanto Kasparov se sentiu frustrado após sua derrota, atribuindo-a, frequentemente, ao descontrole emocional por que passou durante o jogo com Deep Blue. À diferença deste último, podemos imaginar *o que é ser como Kasparov*. Kasparov é consciente e a maioria de suas jogadas foi acompanhada por experiências conscientes – talvez apenas algumas tenham sido executadas "automaticamente". O mais provável é que essas experiências conscientes tenham provocado, no decorrer do jogo, seu descontrole emocional e, finalmente, sua derrota.

Ora, essa comparação entre Deep Blue e Kasparov, proposta por Flanagan, retrata em grande parte por que a ciência cognitiva não se ocupou em tentar explicar a natureza da consciência, até o final dos anos de 1980. E por que, igualmente, quando começaram a surgir teorias da consciência na década de 1990, essas se contentaram em ser teorias da mente, deixando de lado a explicação da especificidade da natureza da experiência consciente.

Para os funcionalistas, Deep Blue é um exemplo de que o processamento inteligente de informação não requer consciência ou de que experiências conscientes são um efeito colateral perfeitamente dispensável na construção de modelos do funcionamento mental. Experiências conscientes podem ser abstraídas na elaboração desses modelos, da mesma maneira que, em teorias físicas, abstrai-se, por exemplo, fatores que possam interferir na consideração do comportamento ideal das moléculas de um gás.

Por outro lado, os naturalistas tenderam a excluir a consideração da experiência consciente nas suas teorias por receio de enfrentar um problema intratável que poderia escapar do domínio de teorias neurocientíficas e que, em última análise, os forçaria a adotar algum tipo de dualismo. Por causa desse receio os naturalistas preferiram desenvolver teorias da mente e não teorias da consciência. Ao equiparar consciência com mente, eles trataram a experiência consciente da mesma maneira que os funcionalistas, ou seja, como uma espécie de efeito colateral ou algo, em última análise, rebarbativo no funcionamento mental.

A estranheza que sentimos diante da afirmação de que Deep Blue seria um modelo de funcionamento da cognição humana não se deve apenas ao fato de essa máquina não ter experiências conscientes e ao reconhecimento de que essas afetam o processamento de informação e o comportamento. Comparemos novamente Deep Blue e Kasparov. Após o jogo, Kasparov ficou deprimido, Deep Blue com certeza não sentiu nada[5]. A inconsciência de Deep Blue torna-o um modelo contra intuitivo da cognição humana na medida em que as experiências conscientes, quando acompanham nossos processos cognitivos, tornam-nos capazes de sentir felicidade, amor, prazer ou dor. É a consciência que torna nossa cognição diferente daquela de um computador ou de insetos que copulam sem sentir prazer. Se quisermos explicar esse tipo especial de cognição, teremos de dispor de uma teoria da consciência – uma teoria que comece por reconhecer a função que essa desempenha em nossa vida mental.

McGinn (1989) chama a atenção para o fato de que é o caráter consciente de nossas experiências o que, em última análise, nos permite ver o filme do mundo em *technicolor*. E chama atenção também para o fato de que o grande mistério que precisamos ainda desvendar é saber como a massa cinzenta de nosso cérebro pode produzir esse filme colorido. Carl Sagan, na sua novela *Contact* [Contato], publicada em 1985, lança um desafio parecido. Ele nos diz: "pense que nesse momento você está num estado consciente. E diga-me se, em algum momento, esse pensamento te remete a algo como alguns bilhões de neurônios disparando nesse instante"[6].

Essas afirmações de McGinn e de Sagan alertam para o risco envolvido nas tentativas de formulação de teorias da consciência. Reconhecer a existência – a ontologia própria – da experiência consciente pode levar-nos ao dilema de não podermos situá-la em nenhum quadro conceitual compatível com uma visão científica do mundo. Por outro lado, rejeitar a existência da consciência pode levar-nos a um empobrecimento teórico inaceitável. Haverá alguma estratégia teórica e metodológica que nos livre desse dilema?

A alternativa que examinaremos é uma nova abordagem à cognição que se desenvolveu na década de 1990, a chamada *neurociência cognitiva*. Já nos referimos a ela,

5. Esta consideração, quase jocosa, deve-se a Flanagan (1998).

6. Cf. Sagan (1985, p. 255).

de passagem, no final do capítulo II. A neurociência cognitiva consolidou-se a partir dos avanços nas técnicas de neuroimagem, permitindo, cada vez mais, uma abordagem empírica da natureza da experiência consciente. Antes, porém, de enveredarmos por uma apresentação dos contornos históricos e metodológicos dessa nova disciplina, examinaremos brevemente algumas teorias da consciência, permitindo ao leitor a compreensão do cenário no qual proliferam teorias e se travam os debates contemporâneos acerca da natureza da consciência.

Uma primeira incursão: Dennett, Calvin e Baars

Caminhando numa direção contrária ao que dissemos acima, Dennett desenvolve uma perspectiva "deflacionária" em relação ao problema da consciência. No seu livro *Consciousness Explained* [Explicando a consciência], publicado em 1991, ele sustenta que o problema da consciência resulta, em grande parte, de falsas percepções que temos de nós mesmos e de nosso próprio funcionamento mental. São essas falsas percepções, frequentemente erigidas em teorias filosóficas que tornam o problema da consciência intratável. O *hard problem* de Chalmers seria um caso típico de pseudoproblema ou de uma mitologia filosófica gerada por tomarmos como ponto de partida para uma teoria da consciência o que nossa cognição sugere que ela é – uma cognição em primeira pessoa, da qual se originam falsas crenças que impossibilitam a formulação de uma teoria empírica da natureza da experiência consciente. Se abandonarmos essa perspectiva, perceberemos que o problema da consciência é muito menor do que ele aparenta ser.

Um dos principais mitos que precisamos derrubar é o *teatro cartesiano*. Supomos que nossas experiências conscientes ocorrem em algum *lugar* na nossa mente, algum tipo de palco interno onde se sucederiam os episódios conscientes que compõem nossa vida mental. Ora, o teatro cartesiano é uma ficção cognitiva, uma metáfora inapropriada resultante de uma falsa concepção de nosso próprio funcionamento mental baseada numa perspectiva de primeira pessoa. No teatro cartesiano entram e saem conteúdos mentais que precisariam, por sua vez, serem transformados em experiências conscientes, numa espécie de segunda transcrição que seria operada por algum tipo de "eu" ou de "self" que funcionaria como um intérprete[7] – um intérprete que por assistir as cenas do teatro daria origem à consciência reflexiva ou autoconsciência.

Dennett move uma crítica severa ao teatro cartesiano, mostrando que dele surgem outros dois mitos correlatos. O primeiro consiste em supor que a esse lugar do teatro cartesiano na nossa mente corresponde algum tipo de correlato neural – um *locus* da consciência que algum dia seria descoberto pela neurociência. Ora, querer descobrir o

7. Este é o "mito da dupla transdução", o que, segundo Dennett, seria a ideia de que todos os dados transmitidos ao cérebro precisariam ser, em algum ponto deste, reinterpretados por "algo" ou "alguém" que dissesse o que eles significam. Cf. Dennett (1996b).

locus da consciência equivaleria a querer descobrir onde está o casamento de um casal quando visitamos sua casa: ele não está em lugar nenhum, embora nessa visita nós percebamos suas consequências o tempo todo. A dívida filosófica de Dennett com as concepções de Ryle torna-se bastante evidente nesse tipo de crítica[8]. Parte dessa tentativa de encontrar os correlatos neurais do teatro cartesiano é achar onde, no corpo, se dá a passagem entre o físico e o mental, isto é, de onde emerge a consciência. Descartes julgava que essa interseção seria a glândula pineal, o que levou Dennett a cunhar esse tipo de tentativa ingênua de alguns neurocientistas de *materialismo cartesiano*. O segundo mito, também correlato ao teatro cartesiano, é supor que exista um único fluxo de consciência, um "significador" central que funcionaria também como intérprete central, que ordenaria as cenas que se passam nesse teatro, tornando-as consistentes e coerentes.

A esse tipo de serialização do fluxo da consciência supostamente produzido pelo teatro cartesiano, Dennett contrapõe o seu modelo de múltiplas camadas (*multiple drafts model*). De acordo com esse modelo, nosso cérebro seria *quase* como uma máquina híbrida ou de arquitetura computacional mista: várias máquinas paralelas acopladas a uma máquina serial. Contudo, essa última seria uma *máquina virtual* produzida pela própria ação desse paralelismo massivo. Vários circuitos especializados no cérebro trabalham em paralelo, realizando diferentes tarefas, criando narrativas fragmentadas, pequenas histórias. Não há um único fluxo de consciência, nem tampouco um "significador" central. É uma ilusão supor que o nosso fluxo de consciência seja unívoco: ele é errático e fragmentário. Em alguns casos essas narrativas são perdidas ou esquecidas, mas outras são mantidas para desempenhar alguma função, por essa máquina virtual no cérebro.

Essa máquina virtual tem um funcionamento serial e gera também uma narrativa serial, mas isto não implica que o funcionamento do cérebro seja serial. A máquina virtual *cria a impressão* de que a narrativa é serial, mas essa é o resultado da competição entre vários fragmentos de narrativas, gerados pelos vários circuitos especializados. Dennett chama essa máquina virtual de *máquina joyceana*, fazendo uma alusão ao *Ulisses* de James Joyce, que retrata o dia de um personagem envolto em episódios de pensamento que caracterizam essa narrativa fragmentária e errática, em uma espécie de ruminação interminável. A máquina joyceana, ao criar a impressão de serialização cria também a ilusão do teatro cartesiano e do "significador" central. Tudo se passa como se a cada segundo houvesse um fragmento vencedor dessa competição e esse fragmento se tornasse então consciente por entrar na narrativa serial, sendo logo em seguida substituído por outro, resultante de uma nova vitória instantânea. É nesse sentido que Dennett nos diz que o que chamamos consciência é um processo que se assemelha à fama, em uma alusão a Andy Warhol. Warhol descreve um mundo imaginário no qual cada ser humano teria o direito a ser famoso por quinze minutos, cedendo, em

8. Cf. novamente o capítulo V. Introduzimos o mesmo exemplo do casamento ao falar de Ryle.

seguida, seu lugar para outro. O mesmo ocorreria com "ser consciente" onde cada fragmento de narrativa entraria na máquina serial por um curtíssimo intervalo de tempo, tornando-se momentaneamente saliente e gerando a sucessão que erroneamente supomos ser uma história coerente acerca de nós mesmos.

Os circuitos especializados que formam a máquina paralela – pelo menos os mais básicos – teriam se originado de processos adaptativos que modelaram nosso cérebro nos primórdios do processo evolucionário, tornando-o apto para resolver problemas específicos e garantindo nossa sobrevivência imediata como organismos num meio ambiente hostil. Posteriormente, esses circuitos teriam sido reaproveitados para executar outras funções distintas das originais. Desse reaproveitamento emerge o dispositivo capaz de alinhar essas narrativas fragmentárias que competem entre si dando origem à máquina joyceana, que surge quando o cérebro começa a processar informação vinda da cultura que está a nossa volta. Nessa está a origem dos conteúdos mentais que são processados na forma de narrativas competitivas. Ou seja, a um determinado estágio do processo evolucionário a atividade cerebral foi submetida a experiências, hábitos de pensamento e dados expressos pela linguagem que invadem os cérebros individuais como se fossem parasitas, transformando-os no que chamamos de mente. As unidades de informação fornecidas pela cultura e processadas pelo cérebro Dennett chama de *memes*, tomando o termo emprestado do biólogo evolucionário R. Dawkins. Memes se espalham pela cultura e pelos cérebros que estão imersos nela da mesma maneira que uma doença infecciosa se espalha num processo epidêmico[9].

A invasão dos cérebros pelos memes e seu processamento pela máquina híbrida que compõem sua arquitetura culmina com a produção de mais uma ilusão que ronda as teorias filosóficas tradicionais da consciência: a ideia de um "ego" ou de um "eu substancial" ao modo cartesiano. O "eu substancial" acaba sendo engendrado na medida em que se forma a ilusão de que haveria uma narrativa privilegiada, a do elaborador central. Mas não é isto o que de fato ocorre: o tempo todo o cérebro está criando inúmeras versões sobre percepções, sensações, emoções, sentimentos. Não há um instante privilegiado para fechamento dessas versões, para se encerrar a edição, nem um circuito ou processador central que faça as vezes de editor-chefe dessa redação. Há apenas uma máquina virtual que entrelaça episódios produzidos pelo *pandemonium* competitivo dos inúmeros circuitos em paralelo.

A formulação de uma teoria da consciência baseada na ideia de uma máquina joyceana ocorreu numa época em que propostas similares estavam sendo desenvolvidas. O neurocientista William Calvin formulara então sua teoria da consciência como resultado de um processo no qual o cérebro funcionaria como uma *máquina darwiniana*.

---

9. Os memes, por analogia com os genes, seriam unidades de transmissão da cultura humana. A hipótese dos memes nunca foi inteiramente aceita como constituindo uma teoria científica, a despeito das tentativas recentes de Lynch (1996), Brodie (1996) e de Blackmore (1999).

No seu livro *The Cerebral Symphony* [A sinfonia cerebral], publicado em 1990, Calvin insistiu na ideia de que estados mentais conscientes advêm de um processo intracerebral essencialmente dinâmico. Não há uma região para a consciência, mas um processo constante no qual alguns conteúdos mentais podem, alternativamente, tornar-se conscientes ou deixar de sê-lo[10].

Para Calvin, a atividade mental tem por finalidade primeira a organização e a orientação do comportamento dos organismos no meio ambiente. Nesse sentido, o cérebro dos organismos representa o meio ambiente para em seguida agir sobre ele. Contudo, o cérebro humano desenvolveu a capacidade de gerar *cenários possíveis* ou representações alternativas do meio ambiente a partir dos dados que recebe, antes de agir. Comportamentos conscientes e comportamentos automáticos são fundamentalmente distintos, mas têm uma raiz comum. O comportamento consciente emerge do comportamento automático quando esse passa a ser precedido de um conjunto de representações ou de cenários possíveis e resulta da escolha de um desses cenários possíveis como guia do curso das ações subsequentes do organismo. A vida mental consciente instaura-se no intervalo entre o recebimento de um input e a produção de um output, pela produção desses cenários possíveis que são causalmente inertes até que um deles seja escolhido para orientar uma ação.

A escolha entre cenários possíveis não pressupõe um intérprete ou um homúnculo no cérebro. Supor a existência de um homúnculo ou intérprete que efetuaria essa escolha é uma ilusão da perspectiva de primeira pessoa. O que chamamos de escolha é um processo de seleção natural intracerebral que ocorre num tempo extremamente acelerado onde os vários cenários competem entre si até que se defina um vencedor. Os vários cenários podem também se combinar entre si, mesclando-se para produzir um novo cenário ganhador que resultaria de um processo semelhante ao de uma mutação entre espécies. É nesse sentido que o cérebro funcionaria como uma autêntica *máquina darwiniana*. Essa ideia percorre a obra de Calvin, tendo sido, posteriormente, aperfeiçoada e até modificada nos trabalhos que sucederam seu livro de 1990, que abordam outros aspectos do funcionamento do cérebro ou a natureza de seus códigos[11].

Além da máquina joyceana de Dennett e da máquina darwiniana de Calvin, as teorias da consciência do início da década de 1990 foram também marcadas pelo aparecimento da teoria do "espaço global de trabalho" (*global workspace*) formulada por Bernard Baars, no seu livro *A Cognitive Theory of Consciousness* [Uma teoria cognitiva da consciência], publicado em 1988. O espaço global de trabalho funciona como uma central de comutação de informações entre os vários processos inconscientes executados por módulos ou circuitos especializados que estão no cérebro. Para executar tarefas rotineiras, o cérebro mobiliza esses módulos, que nos fornecem procedi-

10. Abordo a obra de Calvin, ainda que rapidamente, em Teixeira (1995).

11. Referimo-nos a Calvin (1996a, 1996b).

mentos a serem seguidos. O mesmo não ocorre quando precisamos resolver um problema ou executar uma tarefa para a qual não há procedimentos definidos. Nesse caso, é preciso trocar informações entre os diversos módulos, o que é feito pelo *global workspace*, um espaço no qual a informação dos processos inconscientes é momentaneamente integrada. Dessa integração surgem processos de coordenação e controle, na forma de experiências conscientes.

O *global workspace* remete-nos à metáfora de um teatro, mas não de um teatro cartesiano. O teatro cartesiano tem uma audiência com um único espectador, uma consciência una e imaterial que assistiria a peça e, ao mesmo tempo, controlaria os personagens da mesma maneira que alguém coordena marionetes. Certamente não é desse tipo de teatro que Baars nos fala. Seu *global workspace* não é, tampouco, algum tipo de máquina virtual que produziria a serialização na forma de uma ilusão subjetiva como quer Dennett, mas uma realidade confirmada por evidências neurobiológicas. Uma delas seria fornecida pelo exame de neuroimagens do cérebro que revelam haver áreas que são ativadas quando se desencadeiam processos conscientes, ou seja, quando há produção de informação integrada, o que não ocorre com processos inconscientes. No caso de aprendizagem de novas rotinas ou procedimentos – que exigem a integração de informação – a neuroimagem revela várias áreas sendo ativadas. À medida, porém, que essas rotinas vão se tornando automáticas, a ativação dessas áreas começa a diminuir. Isto significaria, entre outras coisas, que processos conscientes não apenas são detectáveis, como ocorrem em lugares específicos do cérebro.

As teorias formuladas por Dennett, Calvin e Baars ainda contam com muitos adeptos. Os três modelos chamam a atenção para aspectos importantes do funcionamento mental. A teoria das múltiplas camadas de Dennett, a recombinação mutante de cenários possíveis de Calvin e o *global workspace* de Baars revivem um problema que ainda hoje inquieta os neurocientistas: como é possível que um dispositivo com arquitetura paralela, como parece ser o cérebro, possa dar origem ao fluxo serial que supostamente caracteriza a experiência consciente e a linguagem? Que tipo de mecanismo existe no cérebro que lhe permite passar do paralelo para o serial? Quais os mecanismos cerebrais responsáveis pela integração da informação? A busca de uma solução para esse problema já ocupou neurocientistas e psicólogos famosos como, por exemplo, K. Lashley, que, em 1951, publicou um artigo clássico sobre esse tema, "The Serial Order of Behavior" [A ordem serial do comportamento]. Mas nenhuma solução definitiva parece ter sido ainda encontrada. A construção de modelos (ou simulações computacionais) que operem a passagem do paralelo para o serial, integrando informações adequadamente e na mesma velocidade que o cérebro o faz, apresenta-se como um desafio preliminar na busca por uma solução para esse problema. Um desafio que, uma vez superado, precisará ainda ser confirmado no que diz respeito à sua plausibilidade neurobiológica.

Interpretar as teorias de Dennett, Calvin e Baars como constituindo uma explicação da natureza da consciência depende de uma rejeição prévia do *hard problem*. Caso contrário elas devem ser lidas como teorias da mente e não da consciência. Em nenhum momento Dennett explica por que um fragmento de narrativa que con-

quista sua fama efêmera torna-se *necessariamente* uma experiência consciente. O mesmo se aplica ao cenário vencedor de Calvin e ao *global workspace* de Baars. Por que o cenário vencedor tornar-se-ia necessariamente consciente? Por que na integração de informação inconsciente opera-se a passagem para a consciência? Não poderiam esses processos ocorrer sem a produção de consciência, isto é, serem executados por um robô ou por um zumbi?

Esse não parece ser um problema detectável apenas no caso dessas três teorias que acabamos de examinar. As dificuldades envolvidas na formulação e na confirmação de teorias da consciência dependem não apenas de como se define, desde o início, o significado do termo "consciência" como também do tipo de interpretação dos dados que a investigação neurocientífica pode nos fornecer. O risco que corremos, nessas interpretações, é extrapolar as evidências empíricas esquecendo-nos que essas, na maioria das vezes, surgem apenas como confirmações dependentes de hipóteses previamente aceitas. Estaríamos girando em círculos sem o perceber: definimos *a priori* o que é consciência para então buscarmos as evidências empíricas que confirmem tal definição. Evidências que poderiam ser usadas, igualmente, como confirmação de hipóteses acerca de aspectos do funcionamento mental que não implicam na produção de experiências conscientes. Se Wittgenstein estivesse vivo possivelmente afirmaria que não é apenas na psicologia que há métodos experimentais e confusão conceitual – o mesmo se aplica à neurociência. Nessa, corremos sempre o risco de enveredarmos por questões filosóficas e epistemológicas sem percebermos. Nesse domínio, mergulhamos, com frequência, em águas que ainda estão muito turvas, pois a filosofia da neurociência ainda está por ser feita.

A redescoberta do cérebro

Paralelamente às tentativas de explicação da natureza da consciência através de vários tipos de máquinas teóricas, a ciência da mente começou a passar por uma grande transformação. Após duas décadas de hegemonia do modelo computacional da mente, as atenções começaram a se voltar novamente para o papel do cérebro como substrato biológico da cognição e da consciência. Até então a ideia predominante era que a mente seria o software do cérebro, ou que a relação entre psicologia e neurociência seria o mesmo que a relação entre software e hardware respectivamente. Como e onde (em que tipo de substrato físico) o software da mente poderia ser implementado constituía, para os funcionalistas, apenas um detalhe técnico.

Essa situação começou a ser alterada a partir do início da década de 1990. Um movimento que ainda era tímido na década anterior, agora se tornava mais visível: a *neurociência cognitiva*. Essa nova disciplina propunha explicitamente uma reconsideração das bases cerebrais da cognição e da consciência, definindo-se, inicialmente, como o resultado de uma colaboração intensa entre neurociência e ciência cognitiva.

A necessidade de enfatizar o papel e a especificidade do *wetware* (outro termo utilizado pelos neurocientistas para designar o cérebro, por oposição ao hardware dos computadores) nos modelos propostos pela ciência cognitiva já era apontada por pes-

quisadores como Trehub (1991) e Gazzaniga (1995). Este último realizou, durante os anos de 1990, uma série de congressos reunindo pesquisadores de várias áreas na tentativa de definir os contornos interdisciplinares da neurociência cognitiva. Nesses congressos questionava-se que tipo de integração poderia resultar do intercâmbio entre neurociência e ciência cognitiva e se esse implicaria algum tipo de predominância de uma abordagem sobre outra. Havia uma oscilação no modo de definir a neurociência cognitiva; se essa seria uma parte da ciência cognitiva que se ocuparia de modelos inspirados na neurociência ou uma parte da neurociência com uma preocupação voltada para as bases neurobiológicas dos processos cognitivos[12].

É difícil imaginar como poderíamos superar esse tipo de oscilação. A rejeição do funcionalismo e da tese da múltipla instanciação não significava que a neurociência poderia prescindir totalmente de modelos computacionais nem tampouco de modelos matemáticos – o que certamente seria um retrocesso. Mas havia a necessidade de se rediscutir que tipo de papel se deveria atribuir a esses modelos na explicação de fenômenos cognitivos. Uma descrição abstrata desses fenômenos, ou seja, abstraída de sua implementação em cérebros reais não era mais aceitável. Não se aceitava mais, tampouco, uma total independência entre os níveis a partir dos quais esses fenômenos poderiam ser descritos, ou seja, entre psicologia (software) de um lado e neurociência (hardware e implementação) de outro: acreditava-se que ambos deviam se influenciar mutuamente[13].

Apesar das dificuldades expressas nessa oscilação – que ainda perdura – os contornos da neurociência cognitiva tornaram-se, hoje em dia, mais nítidos. Rugg (1997) salienta que a neurociência cognitiva é, antes de mais nada, uma estratégia metodológica, salientando que vários fatores contribuíram para sua formação. Para começar, uma aproximação entre a neuropsicologia clínica e a psicologia cognitiva no estudo dos efeitos das lesões cerebrais, além da observação sistemática das correlações entre comportamentos explícitos de animais e sua atividade neuronal. Esta última tarefa seria executada pela introdução de eletrodos nos cérebros desses animais sem que esses estejam anestesiados. Além disto, a neurociência cognitiva passou a servir-se das novas técnicas de neuroimagem que permitiram, no caso dos seres humanos, o estudo da atividade cerebral *in vivo*.

A neuroimagem é um dos maiores pilares da neurociência cognitiva. Desde a descoberta dos raios-X e do eletroencefalograma (EEG) o aparecimento de novas técnicas, como o PET (Positron Emission Tomography) e o MRI (Magnetic Resonance Imaging) abriram novas portas para que se possa estudar as bases biológicas e cerebrais do comportamento dos cérebros de pessoas vivas, possibilitando uma progressiva integração entre psicologia experimental e neurociência.

---

12. Para uma história mais detalhada da formação da neurociência cognitiva cf. Gazzaniga, Ivry & Mangun (1998), capítulo I. Esta mesma observação me foi feita, em comunicação verbal, pelo Prof. Alfredo Pereira Jr.

13. Cf. Kosslyn e Andersen (1992).

A técnica conhecida como PET baseia-se na possibilidade de "marcar" o oxigênio e a glucose de forma que possamos então "seguir o seu caminho no cérebro". A "marca" é um átomo radioativo que emite pósitrons, ou seja, partículas semelhantes aos elétrons com a diferença de que, ao contrário destes últimos, pósitrons têm uma carga positiva. Injeta-se na veia da pessoa, juntamente com água ou com glucose, átomos radioativos de oxigênio. A marca radioativa segue, então, através da corrente radioativa, até chegar ao cérebro. Os pósitrons chocam-se com os elétrons que estão nas moléculas que se encontram dentro do cérebro até que as cargas positiva (pósitron) e negativa (elétron) se anulem mutuamente. Nesse processo há emissão de grande quantidade de raios gama que atravessam o crânio do indivíduo, podendo, assim, ser detectados por sensores que, por sua vez, podem produzir uma imagem do cérebro em funcionamento, pois a glucose e o oxigênio se acumulam naquelas áreas do tecido cerebral onde a atividade é mais intensa. Através dessa técnica é possível discriminar entre as áreas do cérebro de uma pessoa quando essa executa atividades diferentes, como, por exemplo, ler um texto ou recitá-lo em voz alta.

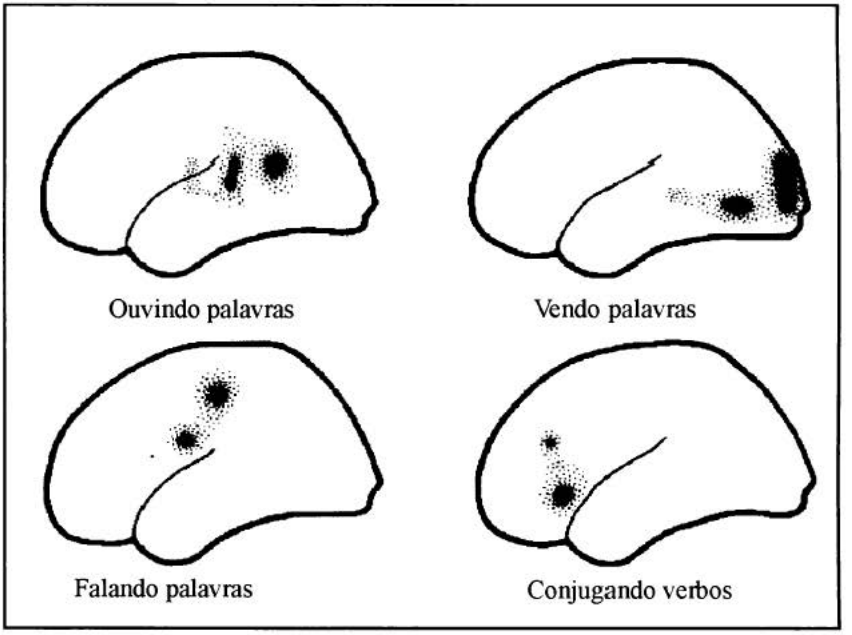

Fig. 7.1 – Ilustração da atividade mental através de neuroimagem. Adaptada de Greenfield (1997).

Um outro tipo de técnica, mais moderna do que o PET, é o MRI. O MRI não precisa de injeções para produzir um contraste. Sua estratégia consiste em medir as mudanças na concentração de oxigênio no sangue que irriga o cérebro. O oxigênio é transmitido pela proteína chamada hemoglobina e o MRI baseia-se no fato de que a

quantidade de oxigênio presente numa área afeta as propriedades magnéticas da hemoglobina. Essas propriedades magnéticas podem então ser monitoradas na presença de um campo magnético onde os núcleos dos átomos se alinham como se fossem magnetos em miniatura. Quando são bombardeados e tirados para fora desse alinhamento através de ondas de rádio, esses átomos emitem sinais de rádio. Esses sinais de rádio expressam a quantidade de oxigênio transportado pela hemoglobina, o que nos permite saber quais as regiões do cérebro que estão mais ativas num determinado momento.

Os resultados que se tem obtido com essas novas técnicas são surpreendentes e inauguram efetivamente uma nova era na neurociência. Além de modificar nossa concepção do funcionamento cerebral, as técnicas de neuroimagem abriram novas perspectivas para o estudo da natureza da consciência. O estudo da atividade cerebral de pacientes com anomalias funcionais como amnésia profunda, prosopagnosia (incapacidade de reconhecimento de rostos), acalculia (incapacidade de efetuar operações aritméticas) e agnosia visual (incapacidade de reconhecimento de objetos) ganhou nova dimensão. Reconhecia-se que explicar a natureza desses distúrbios pressupunha explicar em que sentido eles alteravam aspectos da vida mental consciente de seus portadores. As técnicas de neuroimagem vêm adquirindo importância crescente na medida em que através delas se começa a estabelecer uma conexão entre alterações de consciência e alterações no cérebro. Contudo, o uso de técnicas de neuroimagem não teria se tornado tão importante para o estudo empírico da consciência se essas não estivessem integradas na estratégia metodológica da neurociência cognitiva.

Flanagan (1998) fornece um exemplo para ilustrar em que sentido a neurociência cognitiva constitui uma metodologia para uma investigação empírica da natureza da consciência. Essa metodologia baseia-se, sobretudo, na integração de vários tipos de estratégias que visam correlacionar os níveis psicológicos, comportamentais e neurológicos da investigação da consciência.

O exemplo ilustra uma tentativa de correlacionar alguns tipos de experiência consciente com atividade neural nos macacos rhesus[14]. O ponto de partida é o estudo do fenômeno conhecido como "rivalidade perceptual". Temos casos de "rivalidade perceptual" quando um mesmo estímulo pode produzir duas percepções conflitantes. Casos típicos desse fenômeno são o cubo de Necker e a figura do "pato-coelho" (*duck-rabbit*) apresentadas a seguir.

14. Os parágrafos abaixo foram adaptados de Flanagan (1998).

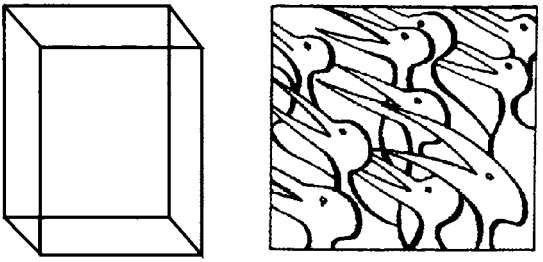

Fig. 7.2 – Ilustrações do cubo de Necker e do *duck-rabbit* (pato-coelho), exemplos clássicos nos quais há ambiguidade perceptual.

Nesses dois casos formam-se, a partir de uma mesma figura, duas percepções possíveis e conflitantes.

Mas há ainda um outro tipo de percepção conflitante, a chamada *rivalidade binocular*. Apresenta-se simultaneamente aos olhos esquerdo e direito dois estímulos visuais incompatíveis. Por exemplo, apresenta-se ao olho esquerdo uma linha subindo e ao direito, uma linha descendo. Experimentos preliminares com percepções incompatíveis mostraram que essas não podem ocorrer simultaneamente para os seres humanos. Nesses casos, o que ocorre é uma alternância entre a percepção da linha se movendo para baixo (olho esquerdo) e da linha se movendo para cima (olho direito).

Suponhamos agora que queiramos saber se esse fenômeno, a rivalidade binocular, ocorre também com os macacos rhesus. Ou seja, queremos saber, através de algum tipo de experimento, se esses macacos têm uma experiência subjetiva semelhante à nossa no caso da rivalidade binocular. O primeiro passo na elaboração desse experimento será treinar o macaco para pressionar uma barra uma vez quando percebe a linha se movendo para baixo e duas vezes quando a percebe se movendo para cima. Pressionar a barra uma vez ou duas vezes funciona como uma espécie de relato que o macaco faz acerca de sua experiência subjetiva. O passo seguinte será correlacionar esse "relato" com eventos no cérebro do macaco. Verificou-se, por exemplo, que há grupos de neurônios que são ativados quando o olho esquerdo recebe estímulos e outros grupos que respondem a estímulos chegando ao olho direito. Há ainda um terceiro grupo que é ativado quando ocorre a mudança de percepção, ou seja, quando a percepção predominante muda do estímulo que chega ao olho esquerdo para aquele que chega ao olho direito.

Esse experimento mostra como a experiência subjetiva pode ser estudada empiricamente através da correlação entre vários níveis de explicação proporcionados por diferentes estratégias teóricas que são integradas pela neurociência cognitiva. Nele se correlacionam experiência subjetiva (experiência visual), comportamento (o macaco foi treinado para fornecer "relatos" de suas experiências através de seu comportamento de pressionar a barra) e a observação de sua atividade cerebral relacionada com a mudança de suas experiências perceptuais.

Essa metodologia integradora – quando aplicada sobretudo a sujeitos humanos com anomalias funcionais – tem como ponto de partida o reconhecimento de uma dimensão própria da experiência consciente e seu papel na cognição. A neurociência cognitiva parte do reconhecimento dessa dimensão para tentar em seguida encontrar suas bases neurais e suas relações com o comportamento. Nesse sentido, seu pressuposto metodológico principal caminha em direção inversa ao reducionismo ou às teorias da identidade, sem entretanto romper com uma proposta materialista. Estratégias identitaristas ou reducionistas, seja quando são adotadas pela neurociência tradicional ou quando são adotadas pela filosofia da mente, tentam explicar a consciência negando a existência e a especificidade desse tipo de fenômeno – na maioria das vezes por não ter podido explicá-lo até agora. Para a neurociência cognitiva, explicar não é necessariamente reduzir. Tampouco teorias da mente poderiam cumprir integralmente o papel de serem teorias da consciência, embora essas últimas envolvam necessariamente a investigação do funcionamento mental e de suas bases neurofisiológicas.

Reconhecer a existência da experiência consciente para em seguida tentar explicá-la significa tomar Kasparov e não Deep Blue como ponto de partida para a construção de um modelo explicativo de como se joga xadrez. Consciência é um tipo de experiência subjetiva. Não sabemos ainda quais são seus correlatos no nível cerebral. Não sabemos se sua produção está relacionada com algum tipo específico de arquitetura neural ou com algum tipo de frequência oscilatória de alguns grupos de neurônios que permite algum tipo especial de codificação de informação. A chave para se responder essas perguntas está no estudo das características da experiência subjetiva e na sua relação com o cérebro e o comportamento.

O fim do funcionalismo?

As tentativas de mapeamento de funções psicológicas em estruturas fisiológicas do cérebro, historicamente, têm causado muitos debates na filosofia da mente e na filosofia da psicologia. A maioria desses debates centra-se no valor explicativo que se deve atribuir a tais mapeamentos. O aparecimento da neurociência cognitiva e sua ênfase crescente no estudo do *wetware* (cérebro), juntamente com a questão da relação entre função e forma neurofisiológica tem contribuído, uma vez mais, para reacender esse tipo de debate. Até que ponto características específicas do cérebro determinam as funções que esse pode realizar? Serão as funções dependentes de formas específicas da arquitetura cerebral?

Nas últimas décadas a análise da noção de função em filosofia da mente tem se mantido sistematicamente atrelada à doutrina funcionalista professada pelos partidários da inteligência artificial. O conceito de múltipla instanciação contribuiu para que a noção de função se mantivesse dissociada de qualquer tipo específico de realidade biológica. Nesse contexto, forma e função são vistos como sendo completamente independentes. Não é o material de que é feito um tabuleiro de xadrez e suas peças nem tampouco seu tamanho ou formato físico que definem esse tipo de jogo, mas a função que lhes é atribuída. Madeira e marfim seriam alternativas físicas válidas a partir das quais se podem construir peças de um jogo de xadrez.

Mas será que o mesmo se aplicaria no caso do funcionamento mental, ou seja, será que esse poderia igualmente ser instanciado num cérebro com características biológicas ou num dispositivo de silício? Haverá alternativas físicas para a atividade eletroquímica do cérebro? Podemos simular um cérebro apenas pela reprodução de suas especificações possíveis como se esse fosse, em sua essência, uma máquina lógica – como defendem os partidários do modelo computacional da mente? Até que ponto as características específicas do material do qual é composto o cérebro determina as funções que esse pode desempenhar?

Os progressos nas tentativas de mapeamento do cérebro têm levado a uma revisão crescente do pressuposto da independência das funções cerebrais em relação às arquiteturas e materiais específicos que a instanciam. Mundale (1998) observa que o delineamento de áreas funcionais distintas no cérebro mostra, cada vez mais, que a psicologia não pode ser isolada da neurociência[15]. O reconhecimento dessa natureza interativa entre esses dois níveis de explicação do funcionamento mental exige uma revisão da proposta funcionalista.

Os pioneiros do funcionalismo, como Putnam e Fodor, sustentam que um mesmo estado mental pode ser produzido por diferentes estados cerebrais e que, inversamente, um mesmo estado neurológico pode produzir vários estados mentais. O que eles não especificam, contudo, é o que devemos entender por um *mesmo* estado mental ou por um *mesmo* estado neurológico. Consideremos, por exemplo, o estado mental "estar com fome". Putnam sustentaria que tanto um ser humano quanto um peixe estariam num mesmo estado mental quando têm fome, apesar de não estarem num mesmo estado neurológico, pois seus sistemas nervosos apresentam grandes diferenças. Mas ambos os estados mentais, do ser humano e do peixe, ao ter fome, seriam funcionalmente equivalentes?

Se considerarmos "estar com fome" a produção e envio de algum sinal para o cérebro do organismo que gere, por sua vez, o desejo de comida, então a fome humana e a fome do peixe podem ser vistas como funcionalmente equivalentes. Contudo, a própria noção de equivalência funcional pode ser questionada. Não dependeria ela de um tipo específico de perspectiva adotada? Uma xícara é funcionalmente equivalente a um copo se os considerarmos a partir da perspectiva de que sua função primeira é "ser um recipiente para beber água". Se atribuirmos à xícara ou ao copo a função de "ser um recipiente para conter água", eles se tornam funcionalmente equivalentes a um regador que também tem a função de "ser um recipiente para conter água". Entretanto, regar um canteiro de flores com um copo ou com uma xícara e não com um regador é algo inadequado. A equivalência funcional envolve uma contextualização que define a atribuição de função.

Considerar a fome do peixe e a fome do ser humano como funcionalmente equivalentes também pressupõe uma contextualização prévia que define a atribuição de fun-

15. Cf. Mundale (1997, 1998). Grande parte do debate exposto nesta seção foi inspirado em Mundale.

ção. A fome do peixe e a fome do ser humano produzem comportamentos distintos nesses dois organismos. No caso do ser humano, ela envolve a preparação da comida ou a ida a um restaurante. O mesmo não ocorre com o peixe. Os estímulos que podem causar fome num ser humano são distintos daqueles que causam fome no peixe. As opções de alimento para um ser humano são também distintas daquelas que podem satisfazer um peixe. Nesse sentido, a fome do ser humano e a fome do peixe só podem ser consideradas funcionalmente equivalentes se consideradas a partir de um contexto específico – um contexto que *abstrai* as suas peculiaridades para torná-las funcionalmente equivalentes.

Ora, esse tipo de abstração teria sido até agora o grande pressuposto da abordagem funcionalista. Um pressuposto que, por levar a ignorar as peculiaridades resultantes dos diferentes tipos de implementação física ou neurológica, estipula, apressadamente, equivalências funcionais entre estados mentais distintos. Estipula também que esses estados mentais podem ser tratados independentemente de qualquer peculiaridade da base física na qual eles podem ser instanciados. Essa teria sido a manobra teórica feita para os funcionalistas proporem o modelo computacional da mente. Com isto, passou-se a ignorar peculiaridades neurológicas ou peculiaridades de forma, que estariam envolvidas na explicação de funções mentais.

O modelo computacional da mente baseia-se na expectativa de que programas computacionais ou neurônios artificiais possam simular os resultados da atividade eletroquímica do cérebro sem que para isto seja necessário *replicar* exatamente a composição biológica e físico-química dos elementos que compõem o tecido cerebral. Com isto ter-se-ia esquecido, por exemplo, que no cérebro há uma variedade de neurotransmissores que produzem efeitos variados, diferenças entre células que executam funções específicas e uma grande variedade de sistemas com suas especificidades. Ter-se-ia esquecido, ademais, que seriam as características físicas do cérebro a chave para explicar como e porquê ele pode desempenhar certas funções.

O resultado dessa grande abstração sobre a qual se apoia o modelo computacional da mente foi abordar a cognição e a consciência como se essas pudessem ser produzidas por *máquinas idealizadas*, desde que essas fossem funcionalmente equivalentes a um sujeito humano. Como consequência, grandes e estéreis debates foram travados entre os filósofos da mente na década de 1970 e 1980; debates quase sempre pueris onde se discutia se máquinas podem ou não pensar, sem se ter sequer uma concepção consensual acerca do que seria o pensamento. Essas discussões foram também, em grande parte, alimentadas por confusões linguísticas resultantes do abuso da metáfora computacional, que levou a uma transposição não só de termos como também de conceitos que passaram a ser empregados indistintamente para designar funções realizadas por mentes e por computadores digitais. "Pensar" é uma delas, "memória" é outra; computadores digitais não têm uma memória e sim um *registro*, embora a memória humana seja também algum tipo de registro.

Defensores mais radicais da primazia do *wetware* sustentariam que o funcionalismo está com os dias contados. Não há dúvida que modelar a cognição ou descobrir os

mecanismos que produzem a consciência através de um conjunto de leis lógicas totalmente independentes do mecanismo físico que as implementa constitui uma estratégia inviável. Essa parece ter sido a lição imediata da neurociência cognitiva e do movimento em direção à redescoberta do cérebro que se iniciou na década de 1990. Será, porém, que a neurociência cognitiva poderá abandonar completamente a utilização de modelos computacionais para estudar o cérebro? A resposta é certamente negativa. Não são os modelos computacionais que devem ser abandonados, mas a pretensão de, a partir deles, podermos construir réplicas completas de atividades cognitivas humanas. Com isto abandona-se um mito que teria sido, por muito tempo, o horizonte implícito da maioria dos defensores do modelo computacional da mente. Contudo, uma nova questão começa a surgir da própria revisão do funcionalismo: serão as características específicas do *wetware* cerebral necessariamente irreplicáveis? Deve o funcionalismo ser abandonado ou simplesmente modificado para que possa incluir, em sua proposta teórica a vinculação de funções psicológicas a características do substrato físico no qual elas são implementadas – tomando a arquitetura do cérebro como ponto de partida?

O futuro da neurociência cognitiva

Dissemos há pouco que as novas técnicas de mapeamento cerebral – sobretudo aquelas que se utilizam de neuroimagem – inauguraram efetivamente uma nova era na neurociência. O estudo aprofundado de características do cérebro levou ao questionamento da proposta funcionalista, sobretudo no que diz respeito à tese da múltipla instanciação. Não se pode mais estudar a mente sem estudar o cérebro, pois se acredita cada vez mais que suas características seriam a chave para a compreensão da natureza da cognição e da consciência. Ora, até que ponto essas novas técnicas de mapeamento utilizadas pela neurociência cognitiva – incluindo, é claro, a neuroimagem – poderiam forçar-nos a concluir que mente e cérebro são a mesma coisa? Estaremos chegando próximos a poder sustentar a veracidade da equação segundo a qual estados mentais *são* estados cerebrais, como defendiam os partidários da teoria da identidade?

A neuroimagem, seja revelando localizações específicas, seja revelando a existência de um sistema integrado, permite-nos apenas inferir a existência de uma *correlação* ou uma *correspondência* entre dois tipos de séries: uma constituída de eventos mentais e outra de eventos cerebrais. Em outras palavras, ela nos permite estipular a existência de um *paralelismo psicofísico*. Será que algum dia, a partir desse paralelismo psicofísico, poderemos esperar que a neuroimagem explique como eventos cerebrais podem gerar estados subjetivos, resolvendo o problema do *explanatory gap*? Essa expectativa torna-se crescente à medida que neurociência e psicologia se aproximam cada vez mais. O grande passo que se espera é que a sofisticação dessas técnicas possa algum dia identificar os correlatos neurais da experiência consciente. Ou seja, que encontraremos alguma característica cerebral que explique a subjetividade ou como estruturas cerebrais podem engendrar a consciência. Um mapeamento cerebral completo e altamente sofisticado representaria muito mais do que estipular correla-

ções entre o mental e o cerebral. Mas, poderão as técnicas de mapeamento extrapolar o paralelismo psicofísico?

Ultrapassar o paralelismo psicofísico remete-nos de volta ao sonho de uma ciência da mente construída em linguagem extensional de que falamos no capítulo III. Contrapusemos as noções de extensão e de intensão (com s) de um termo. Vimos que a extensão de um termo corresponde à classe das coisas às quais esse termo se refere. Vimos também que nem sempre a especificação da extensão de um termo coincide com a especificação de sua intensão (com s). Esses são os casos, por exemplo, de "estrela da manhã" e "estrela da tarde" ou de "Jocasta" e "mãe de Édipo". Nesses casos, especificar extensões não determina o significado ou intensão do termo. Ao discutir os problemas enfrentados pela teoria da identidade mente-cérebro, identificamos como uma de suas principais dificuldades que uma diferença de intensão ou uma diferença de significado não seria escrutável em termos de uma diferença de estado cerebral.

Ora, o mesmo tipo de dificuldade reaparece nas tentativas de se fazer mapeamentos entre estados mentais e estados cerebrais usando técnicas de neuroimagem. Essas são incapazes de detectar diferenças de significado pela identificação de correlatos neurais. Inferir correlatos neurais de estados mentais pressupõe a especificação de significados, se, com essa técnica se deseja obter um mapeamento detalhado e sofisticado. Essa especificação prévia nos devolve a um paralelismo psicofisico e nos força a adotá-lo como condição para a utilização dessa técnica. No caso dos seres humanos, essa especificação de significados pode envolver sua descrição verbal como recurso para torná-los disponíveis para quem vai aplicar a técnica. A dificuldade, contudo, amplia-se no caso de organismos sem habilidades linguísticas, podendo tornar ainda mais difícil o projeto de encontrar correlatos neurais de atividades mentais, pois, nesses casos, significados podem manter-se inescrutáveis.

Para ilustrar essa dificuldade podemos imaginar uma situação na qual se pretende "escanear" o cérebro de uma rã quando essa salta para agarrar com sua língua um inseto qualquer. Escolhemos esse exemplo pelo fato de as bases neuronais desse comportamento da rã já terem sido exaustivamente estudadas por pesquisadores como Lettvin e Maturana, no final dos anos de 1950. Esses estudos mostraram que os olhos desses pequenos animais fazem um recorte seletivo no seu meio ambiente discriminando apenas objetos específicos como pequenos objetos em movimento (insetos, por exemplo) e grandes vultos (predadores). Como resultado dessa atividade de discriminação operada pelo aparelho ótico, são gerados conteúdos perceptuais que ativam uma rede neuronal que, por sua vez, faz com que a rã lance sua língua para fora e tente capturar pequenos objetos voadores sempre que esses aparecem no seu campo visual. Nessa história poderíamos até prescindir da hipótese da formação de conteúdos perceptuais e supor que a discriminação do objeto voador operada pelo aparelho ótico da rã e a ativação de um tipo de comportamento são praticamente uma única e mesma coisa.

Ao "escanearmos" o cérebro da rã quando ela executa esse tipo de comportamento encontraríamos (por hipótese) sempre uma mesma área de seu cérebro sendo ativada.

Mas, poderíamos saber *a que* a rã está reagindo simplesmente pela observação da neuroimagem de seu cérebro? Note-se que a rã reage da mesma forma, sejam esses pequenos objetos em movimento insetos ou pregos que alguém, perversamente poderia jogar na frente do pequeno animal. Essa é, aliás, uma característica que torna esses animais extremamente vulneráveis a qualquer mudança de ambiente.

Temos uma boa margem de pressuposições para sabermos *a que* a rã reage – todas elas baseadas na observação do meio ambiente no qual se situa esse animal. Se no ambiente houver moscas, teremos boas razões para supor que seus olhos estejam detectando moscas. Mas se a rã for transportada para outro ambiente, onde os pequenos objetos em movimento não forem moscas e sim pulgas gigantes e, o mesmo comportamento de saltar e agarrar essas pulgas com a língua se mantiver (e parece que esse é, de fato, o caso), então teremos boas razões para supor que há uma subdeterminação daquilo que a rã pode estar detectando, ou seja, a observação do comportamento do animal e sua correspondente neuroimagem cerebral serão insuficientes para a determinação, seja de um possível conteúdo mental da rã, seja a que seu olho estaria reagindo. Ela estaria reagindo àquilo que nós, observadores de seu comportamento num determinado ambiente, supomos ser o caso, mas aqui encontramos um razoável espectro de possibilidades. Não podemos inferir da neuroimagem um conteúdo mental específico, ou seja, essa é insuficiente para determinar os aspectos intencionais que estariam envolvidos na percepção da rã e nos conteúdos mentais a ela correspondentes. Ou seja, a observação do comportamento da rã e de sua neuroimagem quando ela lança sua língua para fora são insuficientes para especificarmos os aspectos intencionais que estariam envolvidos seja na percepção, seja na discriminação operada pelos olhos da rã. Qualquer inferência da neuroimagem para o conteúdo mental ou a discriminação operados pela rã serão uma construção do observador, ou seja, de quem está aplicando a técnica.

Significa isto que as tentativas de mapeamento cerebral e o projeto da neurociência cognitiva devam ser abandonados? É obvio que não. A conclusão pelo paralelismo psicofísico e pela existência de elementos irredutíveis na relação mente-cérebro não desmerecem esse projeto. Se há uma irredutibilidade, essa está deixando de ser um problema metafísico ou um grande mistério intransponível que nos remeteria para algo sobrenatural, fora da natureza, para tornar-se, gradualmente, o reconhecimento de uma condição metodológica inerente a qualquer tentativa de se construir uma ciência da mente.

O reconhecimento dessa condição metodológica se impõe à medida que, contrariamente às expectativas, o avanço do mapeamento cerebral mostra que apenas algumas funções psicológicas simples podem ser localizadas, o mesmo não ocorrendo com funções mais complexas como a memória, a atenção e a consciência. Tentativas de mapeamento de funções psicológicas mais complexas nos remetem cada vez mais a uma concepção do cérebro como um sistema integrado, no qual não seria possível traçar relações ponto a ponto entre eventos mentais e eventos cerebrais, afastando-nos cada vez mais do reducionismo. Mais do que uma condição metodológica, a impossi-

bilidade da redução da mente ao cérebro parece ser um limite epistemológico imposto pela nossa própria situação cognitiva.

O QUE LER

McGINN, C. *The Problem of Consciousness*

DENNETT, D. *Consciousness Explained*

BAARS, B. *A cognitive Theory of Consciousness*

SHEAR, J. *Explaining Consciousness: the Hard Problem*

FLANAGAN, O. *Consciousness Reconsidered*

BLOCK, N., FLANAGAN, O. & GÜZELDERE, G. (orgs.) *The Nature of Consciousness*

RUGG, M. (orgs.) *Cognitive Neuroscience*

POSNER, M.I. & RAICHLE, M.E. *Images of Mind*

Verbete "Neuroimaging" em BECHTEL, W. & GRAHAM, G. (orgs.) *A Companion to Cognitive Science*

Conclusão

Será que um dia encontraremos uma solução para o problema das relações entre mente e cérebro? Nossas tentativas de compreender o funcionamento cerebral e de relacioná-lo com a produção de nossa vida mental são ainda incipientes e, apesar dos avanços recentes, a ciência cognitiva e a neurociência ainda estão na sua infância. Estamos a anos-luz de distância de replicar a inteligência seja em computadores, seja em robôs. Os problemas a serem enfrentados ainda são gigantescos e, na sua grande maioria, mais conceituais do que propriamente técnicos.

Por outro lado, é inegável que a "década do cérebro" – os anos de 1990 – abriu novas perspectivas para tentarmos resolver esse quebra-cabeças e que, nos últimos anos, a neurociência passou a ocupar um lugar privilegiado no projeto de construir uma ciência da mente. Seria apressado, entretanto, supor que o avanço da neurociência poderia substituir integralmente a abordagem interdisciplinar da mente que se consolidou nas últimas décadas. Modelos computacionais dos correlatos neurais da cognição e da consciência – derivados da ciência cognitiva – continuarão ocupando lugar de destaque nessa investigação. O neurocientista conta com o auxílio do cientista cognitivo para construir simulações computacionais das diversas atividades cerebrais para poder testar – ainda que apenas de forma aproximada – a validade de suas hipóteses.

A "década do cérebro" trouxe uma nova maneira de conceber as relações entre as diversas disciplinas que devem compor uma ciência da mente. Disciplinas como a ciência da computação, a psicologia, a linguística e a filosofia da mente se aproximam cada vez mais da neurociência. A aliança progressiva da ciência da mente com concepções biológicas do funcionamento mental ficara ofuscada pelos exageros do funcionalismo na década de 1970 e sua metáfora da mente como sendo um computador digital.

O risco que corremos agora é o de exagerar no sentido oposto. A concepção do cérebro como uma máquina eletroquímica com propriedades especiais, como a plasticidade e a auto-organização, parece inclinar alguns neurocientistas entusiastas em direção à ideia de que essas propriedades especiais não poderiam ser replicadas artificialmente. Tudo se passaria como se, do fato de o cérebro ter sido até agora um acidente evolucionário único e glorioso, pudesse-se deduzir que nada poderia replicá-lo ou simulá-lo. O funcionalismo deveria então ser abandonado, por pressupor a tese da múltipla instanciação.

Mas essa conclusão parece apoiar-se numa interpretação equivocada da proposta funcionalista. Certamente não somos máquinas idealizadas nem tampouco nossas atividades mentais podem ser replicadas fornecendo-se delas apenas uma *descrição abstrata* na qualidade de um software que poderia ser rodado em qualquer tipo de máquina independentemente de sua arquitetura. Contudo, múltipla instanciação não significa instanciação *irrestrita*. Não é qualquer tipo de substrato físico ou de hardware que pode simular a vida mental – é isto que a neurociência tem procurado nos ensinar. Mas a neurociência não nos ensina que o cérebro é necessariamente irreplicável; tampouco que não podemos reproduzir suas características funcionais usando outros materiais e arquiteturas para simular a mente – da mesma forma que uma máquina de diálise simula um rim.

A múltipla instanciação irrestrita baseia-se na ideia de que haveria uma classe ilimitada de hardwares que poderiam reproduzir o software da mente – a classe das máquinas digitais, que teriam apenas uma característica comum, qual seja, a capacidade de efetuar computações. Esse pressuposto levou a uma falsa equiparação entre a proposta funcionalista e o modelo computacional da mente. No outro extremo, tenta-se derivar da neurociência a ideia de que somente seres dotados de um cérebro semelhante ao nosso poderiam pensar e ter experiências conscientes, como se somente os pássaros pudessem voar e não os aviões, por serem feitos de metal e não terem asas.

Esse debate equivocado entre neurociência e ciência cognitiva que observamos nos últimos anos levou a um afastamento indesejável entre essas disciplinas. Superar essa dispersão é uma das tarefas necessárias para o desenvolvimento de uma ciência da mente no século XXI. Essa tarefa, porém, só será realizada se esse reencontro ocorrer da mesma maneira que um grupo de pessoas se encontrar numa praça, no mesmo dia e na mesma hora, sem ter combinado previamente nenhum compromisso, nem atendido a qualquer tipo de anúncio.

Mas significará esse encontro que a ciência da mente do século XXI poderá resolver todas as questões que nós *queremos* que ela resolva? Ou seja, chegaremos algum dia a uma teoria definitiva acerca da natureza da cognição e da consciência? Ou estará qualquer ciência da mente condenada a uma infância perpétua?

Na primeira parte deste livro percorremos quase todas as possibilidades conceituais que nos permitem conceber algum tipo de relação entre mente e cérebro. Apesar de todas as críticas que fizemos às variedades do materialismo, em nenhum momento afirmamos que esse seria uma hipótese implausível. O materialismo foi o horizonte não só da ciência como também da filosofia da mente que se fez no século XX. O materialista participou do sonho dos físicos de encontrar uma teoria única da natureza que

explicasse toda a diversidade dos fenômenos, incluindo a vida mental e a consciência. É o sonho da Teoria Unificada ou a TOE (*Theory of Everything*)[1].

A TOE ainda é um horizonte remoto, mas a curto prazo é bem provável que a integração entre neurociência e ciência cognitiva torne mais tênue o hiato que hoje existe entre mente e cérebro. Será então que mente e cérebro se tornarão a mesma coisa e que o conceito de mente desaparecerá, à medida que formos encontrando todos os correlatos neurais de nossos fenômenos mentais? Essa foi a grande aposta do século XX: tornar o problema da relação entre mente e cérebro um problema empírico, um problema científico, diminuindo cada vez mais a esfera de especulações e de considerações *a priori* derivadas da filosofia da mente. Mas o sucesso dessa aposta pode nos levar a vários tipos de dificuldades.

Grande parte do debate contemporâneo acerca das relações entre mente e cérebro tem se concentrado numa discussão acerca do tipo de *problema* que enfrentamos. Discute-se se esse é um problema real ou não e como se poderia delimitar seus contornos. Raramente se discute o que seria uma possível *solução* para esse tipo de questão, ou uma solução *desejável*. É possível, por exemplo, que um pequeno conjunto de cientistas chegue a formular uma teoria que identifique o pensamento com algumas reações eletroquímicas que ocorrem no nosso cérebro. Ou que identifique a produção dos *qualia* com algum tipo de fenômeno quântico que ocorreria no cérebro. Essas seriam teorias extraordinariamente complexas e acessíveis apenas para um punhado de especialistas. Não seriam, entretanto, teorias acessíveis ao senso comum, e, por isso, não seriam *soluções desejáveis*: não nos satisfaríamos com elas. Pois a solução que buscamos, ao formular o problema da relação mente-cérebro na qualidade de um quebra-cabeças conceitual, é algo acessível ao senso comum, algo que seja pelo menos *imaginável*.

Uma teoria abrangente do funcionamento do cérebro e de como esse produz uma mente terá de ser necessariamente complexa. Pois essa teoria terá de explicar como pudemos produzir todas as teorias de que dispomos até agora, incluindo teorias complexas como é o caso, por exemplo, da mecânica quântica. E terá de explicar como essa teoria abrangente do funcionamento cerebral pôde produzir a si mesma. Ou seja, ela terá sempre de ser *mais* complexa do que qualquer teoria que possamos produzir. Essa teoria abrangente terá de "herdar o mundo" – e esse seria o autêntico significado de uma TOE. A verdadeira TOE não seria apenas a unificação de todas as teorias existentes, mas a teoria de como o cérebro pode produzir tudo o que está a sua volta, inclusive a si mesmo.

---

1. Utilizo este termo numa acepção modificada em relação ao sentido empregado pela física contemporânea.

Não poderíamos afirmar que a formulação dessa TOE seria impossível, mas podemos questionar se ela seria inteligível. Se essa TOE deve abarcar e basear-se em teorias físicas, as primeiras dificuldades começam a surgir. Por exemplo, é consenso entre os físicos que a mecânica quântica é uma teoria verdadeira, mas que algumas de suas proposições são ininteligíveis, por se afastarem radicalmente de nosso senso comum. Essas seriam proposições *concebíveis*, embora não *inteligíveis*. Ou seja, estamos diante de uma teoria concebível, embora com partes ininteligíveis. Há uma grande diferença entre essas duas qualificações. O velho Descartes já notara nas suas *Meditações* que podemos *conceber* um quilógono – um polígono de mil lados –, mas que não poderíamos *imaginá-lo*. O mesmo ocorreria com a TOE.

No caminho para a formulação dessa TOE defrontamo-nos com o problema do *tipo de conhecimento* que ela seria. Não só a falta de inteligibilidade seria o grande problema, como também a complexidade envolvida na tentativa de descrever a relação entre mente, cérebro e comportamento. Uma complexidade que se inicia ao nível molecular no cérebro, expandindo-se para vários outros níveis até chegarmos aos estados mentais, à experiência consciente e, dando mais um passo, àquilo que chamamos de comportamento. Relacionar esses três níveis e prever o que ocorre quando se passa de um nível para outro seria uma tarefa hercúlea, que envolveria a formulação de uma complexa teoria que teria por objetivo explicar e predizer como se determinam nossos comportamentos. Esse parece ser o sonho daqueles que se empenham em construir uma ciência da mente: diminuir a margem de imprevisibilidade que observamos no comportamento humano; efetuar, a partir dessa ciência da mente, predições que equipariam esta última àquilo que supostamente achamos que a ciência da natureza teria feito. Um raciocínio no qual se esquece que certas disciplinas, como, por exemplo, a meteorologia, não são menos científicas por não acertarem todas as suas previsões, por lidarem com sistemas complexos como é o caso da atmosfera que nos cerca.

Felizmente *não precisamos* de teorias complexas desse tipo, nem mesmo para lidar com regiões do mundo onde a complexidade e a imprevisibilidade predominam, como ocorre com a mente e com o comportamento humano. A natureza e a evolução nos proporcionaram uma teoria muito simples e útil para lidarmos com essas situações de imprevisibilidade e conseguirmos descrições eficientes – embora apenas aproximadas – de nós mesmos e dos seres que nos cercam: a *folk psychology*. Seu principal conceito, "mente", é a ferramenta que usamos para lidar com essa situação, um conceito que expressa naturalmente imprevisibilidade e complexidade. Não atribuiríamos "mentes" a seres cujos comportamentos fossem rígidos e inteiramente previsíveis – é por isso que não queremos atribuir mentes a computadores e robôs, preferindo equipará-los a seres estúpidos numa operação corriqueira na qual identificamos "inteligência" com "mente". Excluímos da esfera dos seres dotados de mente aqueles que não

são capazes de exibir algum tipo de comportamento surpreendente diante de uma situação inesperada. O mais interessante é que sempre utilizamos o conceito de mente muito antes de termos qualquer teoria acerca da imprevisibilidade e complexidade inerentes ao cérebro e ao comportamento humanos. O conceito de mente é algo extremamente primitivo na história da humanidade.

Mesmo que um dia o sonho de um conhecimento completo do cérebro e dos correlatos neuronais de todos os fenômenos mentais se realize, ainda assim o conceito de mente não desapareceria. Substituir a *folk psychology* pelo "neurologuês" (como propõem os partidários do materialismo eliminativo) equivaleria a usar mecânica quântica para projetar pontes e casas. O aparecimento da mecânica quântica e a revisão de nossos conceitos de matéria, espaço e tempo foi incapaz de substituir nossa percepção comum do mundo. A imagem de mundo fornecida pela física clássica – que em boa parte coincide com aquilo que nossa cognição nos fornece – não desapareceu, passando a conviver com o conhecimento fornecido pela física contemporânea. Esse não foi capaz de fornecer – e não poderia – qualquer tipo de imagem de mundo que permitisse orientar nossas ações cotidianas. Essa é, talvez, uma das principais razões pelas quais a imagem clássica de mundo persiste, apesar dos avanços da ciência e das contradições que essa persistência tem produzido.

A razão dessa persistência, no caso da imagem clássica do mundo e igualmente no caso da *folk psychology*, deve-se à extraordinária utilidade de ambas. Intenções, crenças, desejos e outras entidades que compõem a *folk psychology* são verdadeiros atalhos para superarmos a complexidade do cérebro e a imprevisibilidade do comportamento de outras criaturas que povoam nosso meio ambiente. Nesse sentido, a *folk psychology* tem uma grande vantagem prática sobre o "neurologuês" e sobre qualquer TOE futura, embora o aparecimento do conceito de mente que a acompanha tenha originado todas as dificuldades teóricas e conceituais da filosofia da mente. Dificuldades teóricas e conceituais que hoje em dia vêm acompanhadas pela angústia que caracteriza as sociedades ocidentais que passaram a duvidar da dicotomia cartesiana entre mente e cérebro sem, entretanto, dispor de uma teoria aceitável pelo senso comum que explique a natureza da consciência – uma teoria que não precise supor que essa seja um elemento estranho ao mundo que nos rodeia.

Será então a busca pela TOE uma paixão inútil? Em nenhum momento poderíamos afirmar isto. A *folk psychology* não é uma teoria completa e coerente – algo que se assemelharia, estruturalmente, a algum tipo de teoria científica, pois abriga várias contradições e desencontros, a despeito de sua extraordinária robustez. Dela nunca poderíamos esperar algo que a TOE pudesse, no futuro, nos proporcionar: a explicação de uma classe de fenômenos que chamamos de "doenças mentais". Nada desafia mais a integridade da *folk psychology* do que a doença mental entendida como doença

cerebral. Deparar-nos com esse tipo de fenômeno força-nos a reconhecer intuitivamente uma dependência da ideia de mente em relação a algum tipo de base física. A doença cerebral força-nos, igualmente, a conceber algum tipo de passagem entre mente e cérebro – uma passagem que, no limite, desafia a própria existência do *explanatory gap* de que falamos no capítulo III. A busca incessante pela elaboração da TOE poderá abrir o caminho para a descoberta de novas drogas para curar as doenças cerebrais. A eficácia dessas novas drogas servirá de confirmação para as hipóteses da TOE – a despeito dessa continuar ininteligível. Da mesma maneira, a mecânica quântica possibilitou a tecnologia de construção de chips e a consequente revolução informática – apesar de, até hoje, algumas de suas proposições continuarem ininteligíveis.

A *folk psychology* continuará a conviver com a neurociência e com a TOE, *apesar* dessas tornarem concebível sua possível redução a algum tipo de mapeamento cerebral. Sua realidade persistirá enquanto ela for uma ferramenta útil para a detecção de padrões de comportamento – padrões tão reais quanto os centros de gravidade são para os físicos. Ou tão reais quanto o valor do dinheiro é para nós, o que extrapola sua realidade física na forma de notas de papel impresso.

Folk psychology e linguagem amalgamaram-se engendrando o vocabulário mentalista e, com esse, nossa distinção cotidiana entre o físico e o mental. Contudo, seria equivocado supor, como alguns filósofos já o fizeram, que o conceito de mente seria apenas uma construção linguística e que o problema mente-cérebro poderia ser resolvido pela análise da linguagem – uma análise que levaria, inevitavelmente, a sua dissolução na forma de um pseudoproblema. A noção de mente atravessa ortogonalmente a psicologia e nossa vida cotidiana, constituindo-se no conceito operacional mais importante que herdamos através da *folk psychology*. A ideia de mente se manifesta através da linguagem, mas isto não quer dizer que ela seja apenas um artifício linguístico ou uma invenção cultural tardia. Se há uma semelhança de natureza entre linguagem e *folk psychology*, esta se situa no caráter irreversível da instauração de ambas no curso da evolução. Não poderíamos usar a linguagem para falar de algo fora dela sem cair em contradição. A instauração da ideia de mente parece seguir o mesmo movimento. Estamos confinados a um mundo de significações e só as manifestações das doenças cerebrais nos sugere que este mundo tem uma base física anterior à instauração desses significados. Falar dessa base física e explicar nossa vida mental a partir dela sempre nos dará a sensação de um salto mortal – o salto que nos permitiria transpor o *explanatory gap*. Um salto mortal que possivelmente nos devolverá ao lugar de onde saltamos, ou seja, ao mundo dos significados onde nos situamos, pois é a mente que pode produzir uma ciência da mente.

Se a *folk psychology* é a infância da ciência da mente, podemos então afirmar que uma parte dessa ciência sempre viverá uma infância perpétua. Uma infância perpétua

que se caracteriza pela incessante busca de metáforas ou modelos que tornem concebível para o senso comum o que seria a mente, buscando *com o que as mentes se parecem*. É essa busca incessante que forma a história da psicologia – uma história na qual os paradigmas se sucedem uns aos outros e onde dificilmente poderíamos encontrar uma continuidade. Buscamos ainda uma identidade para a mente, equiparando-a seja a um sistema hidráulico, como queria Freud, seja a um computador, como querem os cientistas cognitivos. Resta agora saber qual será a nova metáfora com a qual essa história prosseguirá no novo século.

"A ciência é a infância perpétua da filosofia" (Daniel N. Robinson).

Bibliografia comentada

Os livros e artigos assinalados com um asterisco (*) estão comentados e são especialmente recomendados para aqueles que desejam se aprofundar no estudo da filosofia da mente. Estão incluídos nesta bibliografia também os livros e artigos citados neste trabalho.

*ALEXANDER, I. *Impossible minds.* Londres: Imperial College Press, 1996.

ARMSTRONG, D. *A materialist theory of the mind.* Londres: Routledge, 1968.
[Livro importante para quem quiser se aprofundar nas teorias da identidade mente-cérebro. Reeditado recentemente pela Routledge.]

BAARS, B. *A cognitive theory of consciousness*. Nova York: Cambridge University Press, 1988. O texto está disponível na internet, no seguinte endereço:
http://www.wrightinst.edu/faculty/~baars/book/

BARRETT, P. & GRUBER, H. (orgs.). *Metaphysics, materialism and the evolution of mind*: Early writings of Charles Darwin. Chicago: The University of Chicago Press, 1974.

*BECHTEL, W. & GRAHAM, G. (orgs.). *A companion to cognitive science*. Oxford: Blackwell, 1998.
[Para o estudante de ciência cognitiva, esse é um item bibliográfico obrigatório. Além de uma excelente apresentação histórica do desenvolvimento da ciência cognitiva, contém verbetes extensos sobre os principais tópicos dessa disciplina, o que torna essa obra particularmente abrangente.]

BICALHO, M.R. *Do objeto da psiquiatria:* Uma reflexão através do problema mente-corpo. São Carlos: Universidade Federal de São Carlos, 1997 [Mestrado – Departamento de Filosofia].

*BLACKMORE, S. *The meme machine*. Oxford: Oxford University Press, 1999.
[Livro controverso, constitui a mais recente tentativa de fundamentar cientificamente a teoria dos memes, originalmente proposta por R. Dawkins. Em linguagem clara e acessível.]

BLAKEMORE, C. & GREENFIELD, S. (org.) *Mindwaves*: Thoughts on intelligence, identity and consciousness. Oxford: Basil Blackwell, 1987.

BLOCK, N. Troubles with functionalism. In: SAVAGE, C.W. (org.). *Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press, 1978, v. 9.

BLOCK, N. (org.). *Readings in the philosophy of psychology*. Cambridge, MA: Harvard University Press, 1980, v. 1.
[Coletânea importante, que contém vários artigos sobre o funcionalismo. Foi reeditada em 1983.]

*BLOCK, N.; FLANAGAN, O. & GÜZELDERE, G. (orgs.). *The nature of consciousness*. Cambridge, MA: The MIT Press/Bradford Books, 1997.
[Importante coletânea sobre consciência, contém artigos de Dennett, Flanagan, Block, Jackson, Searle, Chalmers e outros. Particularmente útil para quem quiser se aprofundar nesse tema.]

BODEN, M. *Philosophy of artificial life*. Oxford: Oxford University Press, 1996.

BOGEN, J.E. "On the neurophysiology of consciousness. Part 1 and Part 2". *Consciousness and Cognition*, v. 4, p. 52-62, p. 137-158, 1995.

BORST, C.V. (org.). *The Mind/Brain identity theory*. Londres: The Macmillan Press, 1970.

BRENTANO, F. *Psychologie von empirischen Standpunkt/ Psychology from an empirical standpoint*. Nova York: Humanities Press, 1973. [1925 Translated by A. Pancurello, D. Terrell, L.L. McAlister].

BRODIE, R. *Virus of the mind*: The new science of the meme. Seattle, WA: Integral Press, 1996.

CALVIN, W. *The Cerebral Code*. Cambridge, MA: The MIT Press/Bradford Books, 1996a.

_____. *How brains think*. Nova York: Basic Books, 1996b [edição brasileira: *Como o cérebro pensa*. Rio de Janeiro: Rocco].

_____. *The cerebral symphony*. Nova York: Bantam Books, 1990.

CARVALHO, M.C.M. (org.). *Filosofia analítica no Brasil*. Campinas: Papirus, 1995.

CHALMERS, D.H. *The conscious mind*. Oxford: Oxford University Press, 1996a [edição espanhola: *La mente consciente*. Barcelona: Gedisa].
[Obra mais importante de Chalmers até o momento. Para aqueles que ainda simpatizam com o dualismo, constitui leitura obrigatória. Nele Chalmers apresenta sua teoria do "dualismo naturalista".]

*_____. *On the search for the neural correlate of consciousness*. 1996b.
Em: http://ling.ucsc.edu/~chalmers/papers/ncc.html

_____. "Facing up to the problem of consciousness". *Journal of Consciousness Studies*, v. 2, n. 3, p. 200-219, 1995.

CHAPPELL, V.G. (org.) *The Philosophy of mind*. New Jersey: Englewood Cliffs, 1962.

*CHURCHLAND, P.M. *The engine of reason, the seat of the soul*. Cambridge, MA: The MIT Press, 1995.

_____. *A neurocomputational perspective*. Cambridge, MA: The MIT Press/ Bradford Books, 1992.

_____. *Matter and consciousness*. Cambridge, MA: The MIT Press/Bradford Books, 1984.
[Este livro, reeditado em 1988, constitui uma interessante introdução à filosofia da mente. Aborda o problema mente-cérebro, a metáfora computacional da mente e aspectos interessantes da neurociência e da neurofisiologia.]

*CHURCHLAND, P.S. *Neurophilosophy:* Towards a unified science of mind/brain. Cambridge MA: The MIT Press/Bradford Books, 1986.
[Livro importante para quem quiser se aprofundar no programa teórico do materialismo eliminativo. Os cinco primeiros capítulos apresentam tópicos importantes de neurociência. Os demais discutem o problema mente-cérebro, o dualismo, o funcionalismo e a proposta do materialismo eliminativo.]

COMTE, A. *Cours de philosophie positive*. Paris: J.B. Baillière et Fils, 1869.

COPELAND, J. "The curious case of the Chinese Gym". *Synthese*, v. 95, p. 173-186, 1993.

CORNWELL, J. (org.). *Nature's imagination*. Oxford: Oxford University Press, 1995.

CRICK, F. & KOCH, C. "Towards a neurobiological theory of consciousness". *Seminars in the Neurosciences*, v. 2, p. 263-275, 1990.

CRICK, F. *The astonishing hypothesis*: The scientific search for the soul. Nova York: Simon & Schuster, 1994.

DAHLBOM, B. (org.). *Dennett and his critics*. Oxford: Blackwell, 1993.

DAWKINS, R. *The selfish gene*. Oxford: Oxford University Press, 1976 [edição portuguesa: *O gene egoísta*. Lisboa: Gradiva].

DENNETT, D. *Brainchildren*. Cambridge, MA: The MIT Press, 1998.

_____. *Kinds of minds*. Nova York: Basic Books, 1996a [edição brasileira: *Tipos de mentes*. Rio de Janeiro: Rocco].

_____. *The myth of double transduction*. 1996b. Em http://www.tufts.edu/~dennett/transduc.htm

_____. "The unimagined preposterousness of zombies: commentary on Moody, Flanagan and Polger". *Journal of Consciousness Studies*, v. 2, n. 4, p. 322-326, 1995a.

_____. *Darwin's dangerous idea*. Nova York: Simon & Schuster, 1995b [edição brasileira: *A perigosa ideia de Darwin*. Rio de Janeiro: Rocco].

* _____. *Consciousness explained*. Boston: Little & Brown, 1991.
[Livro pioneiro na exploração de teorias da consciência na filosofia da mente contemporânea. Nele Dennett apresenta sua crítica do teatro cartesiano e propõe o modelo das múltiplas camadas.]

_____. "Real patterns". *Journal of Philosophy*, v. 88, p. 27-51, 1991a.

_____. *Content and consciousness*. Londres: Routledge & Kegan Paul, 1969.

DESCARTES, R. Discours de la Méthode, 1637. In: *Oeuvres philosophiques de Descartes*, presentés par F. Alquié, Tome I. Paris: Garnier Frères, 1963 [edição brasileira: *Discurso do método*. São Paulo: Abril Cultural].

_____. Les Passions de l'ame, 1649. In: *Oeuvres philosophiques de Descartes*, presentés par F. Alquié, Tomo III. Paris: Garnier Frères, 1963 [edição brasileira: *As paixões da alma*. São Paulo: Martins Fontes].

_____. Méditations, 1641. In: *Oeuvres philosophiques de Descartes*, presentés par F. Alquié, Tome I. Paris: Garnier Frères, 1963 [edição brasileira: *Meditações*. São Paulo: Abril Cultural].

DREYFUS, H. (org.). *Husserl, intentionality & cognitive science*. Cambridge, MA: The MIT Press, 1982.

EDELMAN, G.M. *Bright air, brilliant fire*: On the matter of the mind. Nova York: Basic Books, 1992.

_____. *The remembered present*: A biological theory of consciousness. Nova York: Basic Books, 1989.

_____. *Neural darwinism*. Nova York: Basic Books, 1987.

EDELMAN, G.M. & TONONI, G. Neural Darwinism: The brain as a selectional system. In: CORNWELL, J. (org.). *Nature's imagination*. Oxford: Oxford University Press, 1995.

ENGEL, Jr., J. (org.). The mental and the physical. In: FEIGL, H.; SCRIVEN, M. & MAXWELL, G. (orgs.). *Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press, 1958, v. 2.

_____. *Surgical treatment of the epilepsies*. Nova York: Raven Press, 1993.

*FLANAGAN, O. Consciousness. In: BECHTEL, W. & GRAHAM, G. *A companion to cognitive science*. Oxford: Blackwell, 1998.

_____. *Consciousness reconsidered*. Cambridge, MA: The MIT Press, 1992.
[Um dos melhores livros sobre o problema da consciência. Discute temas importantes, como, por exemplo, consciência e cérebro, *qualia*, subjetividade, etc. Defende, de forma coerente, a possibilidade de uma teoria naturalista da consciência.]

FODOR, J. Methodological solipsism considered as a research strategy in cognitive psychology. In: D. *Representations*: Philosophical essays on the foundations of cognitive science. Cambridge, MA: The MIT Press/Bradford Books, 1981.

_____.*The language of thought*. Nova York: Crowell, 1975.

FREUD, S. *Escritos breves 1937-1938*. Buenos Aires: Amorrortu, 1976.

GATES, J.R. et al. Reevaluation of corpus callosotomy. In: ENGEL Jr., J. *Surgical treatment of the epilepsies*. Nova York: Raven Press, 1993.

GAZZANIGA, M.S. *The cognitive neurosciences*. Cambridge, MA: The MIT Press, 1995.

_____. "Right hemisphere language following brain bisection: a 20-year perspective". *American Psychologist,* v. 38, p. 525-537, 1983.

_____. "The split brain in man". *Scientific American,* v. 217, p. 24-29, 1967.

_____. *The cognitive neurosciences*. Cambridge, MA: The MIT Press, 1995.

GAZZANIGA, M.S.; BOGEN, J.E. & SPERRY, R.W. "Some functional effects of sectioning the cerebral commissures in man". *Proceedings of the National Academy of Sciences*, v. 48, p. 1765-1769, 1962.

GAZZANIGA, M.S.; BOGEN, J.E. & SPERRY, R.W. "Observations on visual perception after disconnexion of the cerebral hemispheres in man". *Brain*, v. 88, p. 216-236, 1965.

*GAZZANIGA, M.S.; IVRY, R.B. & MANGUN, G.R. *Cognitive neuroscience*: The biology of the mind. Nova York: W.W. Norton & Company, 1998.
[Excelente livro para quem quiser começar a estudar essa nova disciplina, a neurociência cognitiva. Em linguagem acessível, os autores percorrem temas importantes: os aspectos históricos que levaram à formação da neurociência cognitiva, relações entre neurociência e psicologia, os métodos da neurociência cognitiva, percepção, memória, linguagem e consciência. Ricamente ilustrado, contém também uma série de

entrevistas curtas com autores importantes que se relacionam direta ou indiretamente com essa área de estudos.]

GESCHWIND, N. "The organization of language and the brain". *Science*, v. 170, p. 940-944, 1970.

_____. "Human brain: left-right asymmetries in temporal speech region". *Science*, v. 161, p. 186-187, 1968.

_____. "Disconnexion syndromes in animals and man". *Brain*, v. 88, p. 237-294, 1965.

GOOD, I.J. *The scientist speculates*: An anthology of partly-baked ideas. Londres: Heinemann, 1962.

GREENFIELD, S. *The human brain*: A guided tour. Nova York: Basic Books, 1997. [Livro introdutório, escrito em linguagem simples, percorre os principais tópicos de estudo da neurociência. Imprescindível para o iniciante.]

* _____. *Journey to the centers of the mind.* Nova York: W.H. Freeman, 1995.

GUNDERSON, K. *Mentality and machines*. Nova York: Anchor Books, 1971.

*GUTTENPLAN, S. (org.). *A companion to the philosophy of mind*. Oxford: Blackwell, 1994.
[Livro fundamental para os interessados em filosofia da mente e ciência cognitiva. A primeira parte contém um excelente apanhado dos principais temas abordados pela filosofia da mente. Os verbetes são escritos pelos melhores especialistas nessas áreas, destacando-se, dentre eles, Bechtel, Block, P. Chuchland, Copeland, Dennett, Fodor, Kim, McGinn, Putnam e Searle – para citar apenas alguns.]

HALDANE, J.B.S. *The inequality of man*. Londres: Chatto & Windus, 1932.

HAMEROFF, S.R. "Quantum coherence in microtubules: A neural basis for emergent consciousness?" *Journal of Consciousness Studies*, v. 1, p. 91-118, 1994.

HANNAN, B. *Subjectivity and reduction*: An introduction to the mind-body problem. Boulder, CO: Westview Press, 1994.

HANSEN, F.C. *Inteligência artificial e o problema mente-corpo*. São Carlos: Universidade Federal de São Carlos, 1995 [Dissertação de mestrado – Departamento de Filosofia].

HAUGELAND, J. *Mind design*. Cambridge, MA: The MIT Press/Bradford Books, 1981.

*HEIL, J. *Philosophy of mind*: A contemporary introduction. Londres: Routledge, 1998. [Uma das mais recentes e atualizadas introduções à filosofia da mente escrita em lín-

gua inglesa. Aborda o cartesianismo, o materialismo, o funcionalismo e o materialismo eliminativo em linguagem clara e acessível.]

HOFSTADTER, D. & DENNETT, D. *The mind's I*. Sussex: The Harvester Press, 1981.

HORGAN, J. *The end of science*: Facing the limits of knowledge in the twilight of the scientific age. Nova York: Addison Wesley, 1996.

HUSSERL, E. *Cartesianische Meditationen und Pariser Vorträge/ Méditations Cartésiennes*. Paris: Librairie Philosophique J. Vrin, 1969. [Traduit de l'allemand par G. Peiffer e E. Levinas.]

JACKENDOFF, R. *Consciousness and the computational mind*. Cambridge, MA: The MIT Press, 1987.

JACKSON, F. "What Mary didn't know". *Journal of Philosophy*, v. 83, n. 5, p. 291-295, 1986.

_____. "Epiphenomenal Qualia". *Philosophical Quarterly*, v. 32, n. 127, p. 127-136, 1982.

JAMES, W. *The principles of psychology*. Nova York: Henry Holt, 1890 [Reimpresso em 1950 por Dover Books].

JEFFRESS, L.A. *Cerebral mechanism in behavior*. Nova York: J. Wiley and Sons, 1951.

KANT, I. *Kritik der reinen Vernunft*, 1781 [*Crítica da razão pura* – Traduzido do alemão por V. Rohden e U. Mossburger. São Paulo: Abril Cultural, 1980].

KIM, J. *Philosophy of mind*. Boulder, CO: Westview Press, 1996.

_____. *Supervenience and mind*. Cambridge: Cambridge University Press, 1993.

KOCH, C.E. & DAVIS, J.L. *Large-scale neuronal theories of the brain*. Cambridge, MA: The MIT Press, 1994.

KOSSLYN, S. & ANDERSEN, R. (orgs.) *Frontiers in cognitive neuroscience*. Cambridge, MA: The MIT Press, 1992.

LASHLEY, K. The problem of serial order in behavior. In: JEFFRESS, L.A. *Cerebral mechanism in behavior*. Nova York: J. Wiley and Sons, 1951.

LETTVIN, J. & MATURANA, H. "What the frog's eye tells the frog's brain". *Proceedings of the Institute of Radi Engineers*, v. 47, n. 11, p. 1940-51, 1959.

LEVINE, J. "Materialism and qualia: the explanatory gap". *Pacific Philosophical Quarterly*, v. 64, p. 354-361, 1983.

LIBET, B. The neural time factor in conscious and unconscious events. In: *Experimental and theoretical studies of consciousness*. Nova York: Wiley, 1993 [Ciba Foundation Symposium, 174].

LLINAS, R.R. & PARE, D. "Of dreaming and wakefulness". *Neuroscience*, v. 44, p. 521-535, 1991.

LLINAS, R.R.; RIBARY, U.; JOLIOT, M. & WANG, X.-J. Content and context in temporal thalamocortical binding. In: BUZSAKI, G.; LLINAS, R.R. & SINGER, W. (orgs.). *Temporal coding in the brain*. Berlim: Springer Verlag, 1994.

LOCKWOOD, M. *Mind, brain and the quantum*. Oxford: Basil Blackwell, 1989.

LOGOTHETIS, N. & SCHALL, J. "Neuronal correlates of subjective visual perception". *Science*, v. 245, p. 761-763, 1989.

LOSANO, M. *Histórias de autômatos*. São Paulo: Companhia das Letras, 1990.

LUTSEP, H.L. "Cerebral and callosal organization in a right hemisphere dominant split-brain patient". *Journal of Neurology, Neurosurgery and Psychiatry,* v. 59, n. 1 p. 50-54, 1995.

LYCAN, W.G. *Mind and cognition*. Oxford: Basil Blackwell, 1990.
[Importante e ambrangente coletânea sobre os fundamentos filosóficos da ciência cognitiva. Inclui também textos importantes de filosofia da mente. Entre os temas abordados estão: o funcionalismo, o materialismo eliminativo, a hipótese da linguagem do pensamento, a *folk psychology* e o problema da consciência e dos *qualia*.]

LYNCH, A. Thought contagion: How belief spreads through society. Nova York: Basic Books, 1996.

MacKAY, D. Divided brains – Divided minds? In: BLAKEMORE, C. & GREENFIELD, S. *Mindwaves*: Thoughts on intelligence, identity and consciousness. Oxford: Basil Blackwell, 1987.

MARKS, C.E. *Comissurotomy, consciousness and the unity of mind*. Cambridge, MA: The MIT Press, 1981.

McCAULEY, R.N. (org.). *The Churchlands and their critics*. Oxford: Blackwell, 1996.

*McGINN, C. *The character of mind*. Oxford: Oxford University Press, 1982.
[Este livro, republicado em 1997, constitui uma instigante introdução à filosofia da mente. Nele são abordados os principais problemas dessa disciplina: a natureza do mental, o problema das relações entre mente e cérebro, as relações entre pensamento e linguagem, etc. Especialmente recomendado para o iniciante.]

McGINN, C. "Consciousness and space". *Journal of Consciousness Studies*, v. 2, n. 3, p. 220-230, 1995.

_____. *The problem of consciousness*. Oxford: Blackwell, 1991.

_____. "Can we solve the mind-body problem?" *Mind*, v. 98, p. 349-366, 1989.

MOROWITZ, H.J. "Rediscovering the mind". *Psychology Today*, v. 14, n. 3, p. 8-12, 1980.

MYERS, J.J. "Right hemisphere language: science or fiction?" *American Psychologist*, v. 54, p. 315-320, 1984.

MUNDALE, J. Brain mapping. In: BECHTER, W. & GRAHAM, G. A companion to cognitive science. Oxford: Blackwell, 1998.

_____. *How do you know a brain area when you see one?* A philosophical approach to the problem of mapping the brain and its implications for the philosophy of mind and cognitive science. St. Louis, MO: Washington University Press, 1997.

NAGEL, Th.* *The view from nowhere*. Oxford: Oxford University Press, 1986. [Este livro é, como o próprio autor diz, "reacionário". Nele Nagel faz uma defesa explícita do dualismo, contrariando todas as tendências da filosofia da mente do século XX. Há que se ressaltar, contudo, que o livro é extraordinariamente bem escrito e chega a ser quase convincente. Reeditado em 1989.]

_____. "What is it like to be a bat?" *Philosophical Review*, v. 83, p. 435-450, 1974.

_____. "Brain bissection and the unity of consciousness". *Synthese*, v. 22, p. 396-413, 1971.

_____. *"Physicalism The Philosophical Review"*, v. 74, n. 3, p. 339-356, 1965.

PENROSE, R. *The emperor's new mind:* concerning computers, minds and the laws of physics. Oxford: Oxford University Press, 1989 [edição brasileira: *A mente nova do rei*. Rio de Janeiro: Campus].

PINEL, Ph. *Traité médico-philosophique sur l'aliénation mentale, ou La manie*. Paris: Richard, Caille et Ravier, 1801.

PINKER, S. *How the mind works*. Nova York: W.W. Norton & Company, 1997 [edição brasileira: *Como a mente funciona*. São Paulo: Companhia das Letras].

PLACE, U.T. Is consciousness a brain process? In: BORST, C.V. (org.). *The Mind/ Brain identity theory*. Londres: The Macmillan Press, 1970.

POPPER, K. & ECCLES, J. *The self and its brain*. Berlim: Springer International, 1977 [edição brasileira: *O eu e seu cérebro*. Campinas: Papirus].

POSNER, M.I. & RAICHLE, M.E. *Images of mind*. Nova York: Scientific American Library, 1994.

PUTNAM, H. Minds and machines. In: PUTNAM, H. (org.). *Mind, language and reality*. Cambridge: Cambridge University Press, 1975, v. 2.

_____. *Mind, language and reality*. Cambridge: Cambridge University Press, 1975, v. 2.

_____."Reductionism and the nature of explanation". *Cognition*, v. 2, p. 131-146, 1973.

REY, G. *Contemporary philosophy of mind*. Oxford: Blackwell, 1997.

ROBINSON, D. *Introdução analítica à neuropsicologia*. São Paulo: EPU, 1973.

ROBINSON, H. (org.). *Objections to physicalism*. Oxford: Clarendon Press, 1993.

RORTY, R. *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press, 1979.

_____. "Mind-body identity, privacy and categories". *The Review of Metaphysics*, v. 19, n. 1, p. 24-54, 1965.

*ROSENTHAL, D.M. (org.). *The nature of mind*. Nova York: Oxford University Press, 1991.

_____. *Materialism and the mind body problem*. Nova Jersey: Prentice Hall, 1971. [Este livro é uma importante coletânea de artigos sobre o problema mente-cérebro. Contém seleções de textos de filósofos clássicos, como Descartes, Spinoza, Hobbes além de textos contemporâneos de J.J.C. Smart, J. Kim, J. Fodor, H. Putnam e T. Nagel. Está sendo reeditado.]

*RUGG, M. (org.). *Cognitive neuroscience*. Cambridge, MA: The MIT Press, 1997. [Importante coletânea de artigos sobre temas importantes da neurociência cognitiva. Inclui trabalhos sobre aprendizado, memória, neuroimagem, etc. dos quais o mais sugestivo é o último, onde S. Köhler e M. Moscovitch discorrem sobre aspectos do processamento visual inconsciente em casos de agnosia visual, acromatopsia, prosopagnosia, e discutem modelos de consciência.]

RUSSELL, B. *The analysis of matter*. Londres: Kegan Paul, 1927.

_____. *The ABC of relativity*. Londres: George Allen & Unwin, 1925 [edição brasileira: *O ABC da relatividade*. Rio de Janeiro: Zahar Editores].

RYLE, G. *The concept of mind*. Nova York: Barnes & Noble, 1949.

SAGAN, C. *Contact: A novel*. Nova York: Simon & Schuster, 1985.

SCHANK, R.C. & ABELSON, R.P. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum, 1977.

SCHOEMAKER, S. "Functionalism and qualia". *Philosophical Studies*, v. 27, p. 271-315, 1975.

SCHOEMAKER, S. & SWINBURNE, R. *Personal identity*. Oxford: Basil Blackwell, 1984.

SCHRÖDINGER, E. Mind and matter. In: *What is life? and Mind and matter*. Cambridge: Cambridge University Press, 1958 [edição brasileira: *O que é vida?* São Paulo: Editora da Unesp].

_____. *Science and humanism*. Cambridge: Cambridge University Press, 1951 [ediçãopor J. Searle sobre o problema da consciência nos quais ele analisa as teorias de F. Crick, G. Edelman, D. Dennett e D. Chalmers. Na edição brasileira merece destaque a apresentação feita pelo Prof. Bento Prado Jr.]

_____. The rediscovery of mind. Cambridge, MA: The MIT Press, 1992 [edição brasileira: *A redescoberta da mente*. São Paulo: Martins Fontes].

_____. *Intentionality*: An essay in the philosophy of mind. Cambridge: Cambridge University Press, 1983 [edição brasileira: *Intencionalidade*. São Paulo: Martins Fontes].

_____. What is an intentional state? In: DREYFUS, H. *Husserl, intentionality & cognitive science*. Cambridge, MA: The MIT Press, 1982.

_____. "Minds, brains and programs". *Behavioral and Brain Sciences*, v. 3, p. 417-424, 1980a.SEARLE, J. Minds, brains and programs. *Behavioral and Brain Sciences*, v. 3, p. 417-424, 1980a.

_____. "Intrinsic intentionality". *Behavioural and Brain Sciences*, v. 3, p. 307-309, 1980b.

SELLARS, W. *Science, perception and reality*. Atascadero, CA: The Ridgeview Publishing, 1991. portuguesa: *A natureza e os gregos e ciência e humanismo*. Lisboa: Edições 70].

* SEARLE, J. *The mystery of consciousness*. Nova York: Nyrev, 1997 [edição brasileira: *O mistério da consciência*. São Paulo: Paz e Terra].
[Apesar de tendencioso e escrito em tom provocador, contém artigos interessantes escritos

_____. *Empiricism and the philosophy of mind*, 1963 [Reimpresso em SELLARS, W. *Science, perception and reality*. Atascadero, CA: The Ridgeview Publishing, 1991].

SHANNON, C. "The mathematical theory of communication". *Bell Systems Technical Journal,* v. 27, p. 379-423, 1948.

*SHEAR, J. *Explaining consciousness*: The hard problem. Cambridge, MA: The MIT Press/Bradford Books, 1995.
[Importante coletânea de artigos sobre o problema da consciência na filosofia da mente contemporânea. Inclui textos de D. Chalmers, D. Dennett, P.S. Churchland, C. McGinn e F. Varela.]

SMART, J.J.C. Sensations and brain processes. In: CHAPPELL, V.G. (org.). *The Philosophy of Mind*. Nova Jersey: Englewood Cliffs, 1962. [Reimpresso também em BORST, C.V. (org.). *The Mind/Brain identity theory*. Londres: The Macmillan Press, 1979].

SMITH, A.D. Non-reductive physicalism? In: ROBINSON, H. (org.). *Objections to physicalism*. Oxford: Clarendon Press, 1993.

SPERRY, R.W. "Hemisphere decconection and unity in conscious awareness". *American Psychologist*, v. 23, p. 723-733, 1968.

_____. "The great cerebral commissure". *Scientific American*, v. 210, p. 42-52, 1964.

SPERRY, R.W. et al. *Psychiatric case formulations*. Washington DC: American Psychiatric Press, 1992.

SWINBURNE, R. Personal identity: The dualist theory. In: SCHOEMAKER, S. & SWINBURNE, R. *Personal Identity*. Oxford: Basil Blackwell, 1984.

TEIXEIRA, J.F. *Filosofia da mente e inteligência artificial*. Campinas: Edições CLE-Unicamp, 1996.

_____. Teorias cognitivistas da consciência. In: CARVALHO, M.C.M. (org.). *Filosofia analítica no Brasil*. Campinas: Papirus, 1995.

TEIXEIRA, J.F. (org.) *Mentes e máquinas*: Uma introdução à ciência cognitiva. Porto Alegre: Editora Artes Médicas Sul, 1998a.

_____. "The allure of brain science". *Ciência e cultura*, v. 50, n. 2, p. 141-145, 1998b [número especial sobre consciência].

_____. "A teoria da consciência de David Chalmers". *Psicologia-USP*, v. 8, n. 2, p. 109-128, 1997 [número especial sobre consciência].

_____. *Cérebros, máquinas e consciência*. São Carlos: Edufscar (Editora da Universidade Federal de São Carlos), 1996a.

TREHUB, A. *The cognitive brain*. Cambridge, MA: The MIT Press, 1991.

WIGNER, E. Remarks on the mind-body question. In: GOOD, I.J. *The scientist speculates*: An anthology of partly baked ideas. Londres: Heinemann, 1962.

WILLIAMS, B. *Descartes*: The project of pure enquiry. Harmondsworth: Penguin Books, 1978.

WITTGENSTEIN, L. *The blue and brown book*. Oxford: Basil Blackwell, 1969.

_____. *Philosophische Untersuchungen/Philosophical investigations*. Oxford: Basil Blackwell, 1951, [Translated by G.E.M. Anscombe].

ZUBOFF, A. The story of a brain. In: HOFSTADTER, D. & DENNETT, D. *The mind's I*. Sussex: The Harvester Press, 1981.